National Research Council Canada
Conseil national de recherches Canada

Division of Physics
Division de physique

SEMINAR ON COLORIMETRY

CONFÉRENCE SUR LA COLORIMÉTRIE

Summaries of Lectures / Résumés des conférences

November 5 - 7 Novembre 1986
OTTAWA

Canada

89-03244/

Lecture Summaries:
NRC Seminar on Colorimetry

November, 1986

National Research Council of Canada
Division of Physics
Ottawa, Ontario, Canada, K1A 0R6

NRCC Publication Number 26489

Preface

This booklet contains summaries of the lectures given at a Seminar on Colorimetry held by the Division of Physics of the National Research Council of Canada on 5-7 November 1986.

The Seminar was one of a series on the measurement of basic physical quantities. The National Research Council holds these Seminars on a regular basis and, as a result, a subject is repeated in time intervals of roughly three to four years depending on interest. Changes in measurement techniques take place during these intervals and it is hoped that a continuing updating of information and a broadening of knowledge in measurement techniques will take place in Canadian industry.

An essential part of the Seminars on Colorimetry is an exhibition of the latest instrumentation. The National Research Council is grateful to the instrument manufacturers who agreed to display their equipment and to contribute to the Seminar.

This booklet was produced with the cooperation of the University of Waterloo Department of Computer Science to whom we are most grateful.

A.R. Robertson,
Head,
Photometry and Radiometry Section,
National Research Council.

Table of Contents

Lecture 1: Physics of Light and Colour 1
Lecture 2: Basics of Human Colour Vision 11
Lecture 3: Basics of Colorimetry 31
Lecture 4: The CIE System 47
Lecture 5: Colour Appearance 59
Lecture 6: Colour Order Systems 67
Lecture 7: Principles of Photometry and Radiometry 79
Lecture 8: Spectroradiometry 89
Lecture 9: Spectrophotometry 101
Lecture 10: Reflectance Standards 115
Lecture 11: Tristimulus Colorimetry 121
Lecture 12: Colorimetry of Video Displays 125
Lecture 13: Gloss Measurements 133
Lecture 14: Technology Transfer 139
Lecture 15: Colour-difference Evaluation 143
Lecture 16: Color Image Reproduction 149
Lecture 17: Applied Colorimetry 157

Lecture 1: A. R. Robertson

THE PHYSICS OF LIGHT AND COLOUR

1.1 Introduction

This seminar is about the measurement and specification of colour. Throughout the seminar it will be important to distinguish between two uses of the word 'colour'. The first use is to describe a physical stimulus that enters our eyes. The second use is to describe the sensation that the stimulus evokes. Physical instruments such as colorimeters, spectrophotometers and spectroradiometers can measure a colour stimulus, but we need to know something about the functioning of the human visual system before we can calculate numbers that adequately describe the colour sensation that the stimulus produces.

This first lecture will be concerned only with the stimulus. It will include a discussion of the stimulus itself, of sources and detectors of colour stimuli, and of processes (reflection, transmission and fluorescence) by which colour stimuli are modified.

Wyszecki and Stiles (1982) is a good source for more detail from a colorimetrist's point of view. Kaufman (1982) gives a lot of information on sources. Budde (1983) gives a thorough treatment of detectors.

1.2 Light

Light is radiant energy, propagated in the form of electromagnetic waves, that is capable of exciting the retina of the eye and producing a visual sensation.

1.2.1 Electromagnetic waves

Electromagnetic waves are oscillating electric and magnetic fields (see figure 1.1) travelling at a speed of approximately 3×10^8 metres per second ($\mathrm{m \cdot s^{-1}}$). The 'speed of light' is always given the symbol c.

For convenience of reference, the electromagnetic spectrum is divided into regions as shown in table 1.1. The division is arbitrary as there is a gradual transition from region to region. The term 'optical radiation' is an additional term sometimes used to describe ultraviolet, visible and infrared radiation.

Wavelength Range	Description
less than 0.1 nm	Gamma rays
0.1 nm to 10 nm	X rays
10 nm to 400 nm	Ultraviolet
400 nm to 700 nm	Visible
700 nm to 1 mm	Infrared
1 mm to 100 km	Radio waves

Table 1.1. Regions of the electromagnetic spectrum.

1.2.2 Photons

Electromagnetic waves travel in discrete bundles that can be thought of as particles, called photons. Each photon has an energy of $h\nu$ where $h = 6.626196 \times 10^{-34}$joule · seconds and ν is the frequency of oscillation. The constant, h, is known as Planck's constant. The frequency, ν , is related to the wavelength, λ, by $\nu = c\lambda^{-1}$. Waves are also sometimes described by the wavenumber, ν', defined by $\nu' = \lambda^{-1}$.

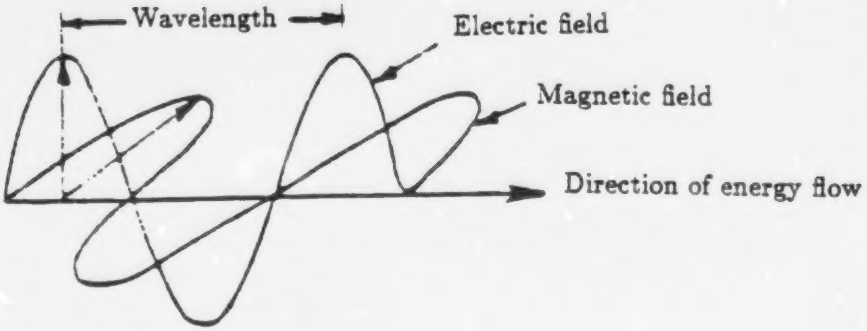

Figure 1.1. Simplified picture of an electromagnetic wave. The oscillations are perpendicular to the direction of energy flow.

1.2.3 Human eye

The human eye is sensitive only to a very small range of wavelengths within the electromagnetic spectrum. The peak sensitivity is for radiation of wavelength 555 nm. The sensitivity falls off for wavelengths higher and lower than this, reaching almost zero at 400 nm and 700 nm (see figure 1.2).

1.2.4 Spectral distributions

To specify a colour stimulus we have to specify the spectral energy distribution. This is the amount of energy at each wavelength between 400 nm and 700 nm. In some special cases the limits are extended to 360 nm and 830 nm although the sensitivity of the eye to these extreme wavelengths is very low. In practice, the specification is simplified by dividing the spectrum into narrow bands (usually 2, 5 or 10 nm wide) and specifying the total energy in each band.

Spectral energy distributions are measured in units such as joules per nanometre. More often we are interested in the rate of flow of energy, or spectral power distribution, which is measured in watts per nanometre ($\mathrm{W \cdot nm^{-1}}$).

In some applications it is more useful to specify the numbers of photons at each wavelength. The transformation is $N = Q/h\nu$ or $N = Q\lambda/hc$ where N is the number of photons and Q is the total energy.

1.2.5 Perceived colour

The perceived colour of electromagnetic radiation varies with the wavelength. Table 1.2 shows the approximate correlation of colour with wavelength.

Most stimuli seen in everyday life contain radiation of various wavelengths in different proportions creating a multitude of different shades and mixtures of the principle colours. By some estimates, the human eye can distinguish 10 million different colours.

Stimuli containing all the visible wavelengths in roughly equal proportions appear white. This can be demonstrated by passing white light through a prism which spreads out the constituent wavelengths, as shown in figure 1.3.

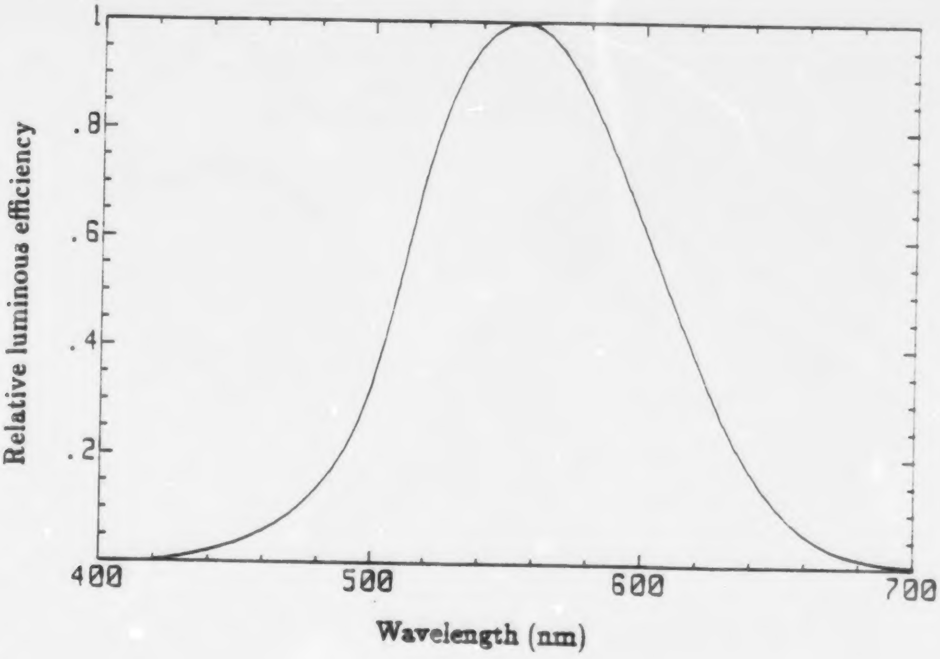

Figure 1.2. Relative luminous efficiency of the human eye.

Wavelength Range	Colour
400 – 430 nm	Violet
430 – 480 nm	Blue
480 – 560 nm	Green
560 – 590 nm	Yellow
590 – 620 nm	Orange
620 – 700 nm	Red

Table 1.2. Dependence of perceived colour on wavelength.

1.3 Sources of Light

1.3.1 Blackbody radiation

A perfectly black object will, when heated, emit more radiant power at each wavelength than any other object at the same temperature. The amount of power can be calculated by a formula known as Planck's Law:

$$M_{e\lambda} = c_1 \lambda^{-5} \left(e^{c_2/\lambda T} - 1 \right)^{-1} \mathrm{W \cdot m^{-3}},$$

where $M_{e\lambda}$ is the power emitted per unit area per unit wavelength interval and T is the temperature in kelvins (K). The name given to $M_{e\lambda}$ is spectral radiant exitance. The constants c_1 and c_2 are known as Planck's radiation constants and have the values:

$$c_1 = 3.7415 \times 10^{-16}\ \mathrm{W \cdot m^2}$$

$$c_2 = 1.4388 \times 10^{-2}\ \mathrm{m \cdot K}$$

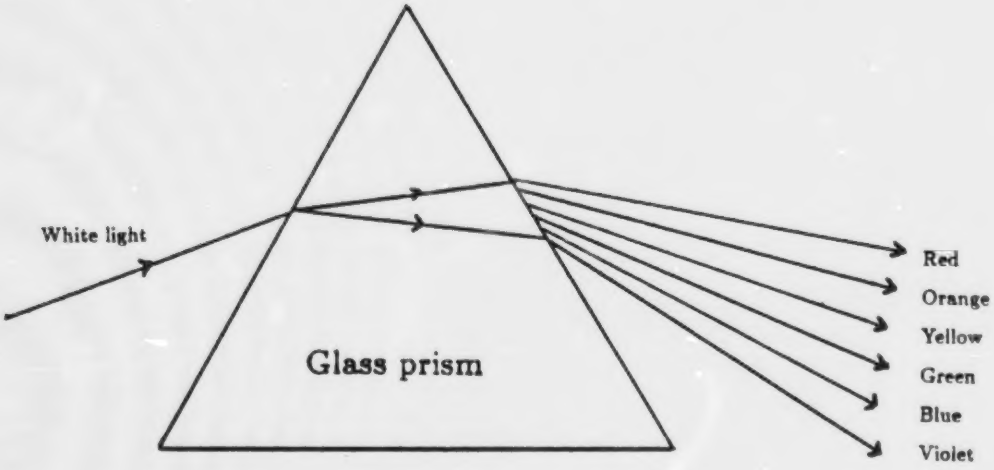

Figure 1.3. Use of a prism to disperse white light into a spectrum.

Alternative names for this series of radiators are 'blackbody radiators', 'full radiators' or 'Planckian radiators'.

The spectral power distributions of some blackbody radiators at different temperatures are shown in figure 1.4. Notice that as the temperature is increased, the colour changes from red through orange and yellow to white and finally to blue.

A good laboratory implementation of a blackbody is a hole in the wall of an otherwise closed chamber. Until 1979 a blackbody like this, heated to the temperature of melting platinum, was used as the international reference standard for maintaining the basic unit of light, the candela.

1.3.2 Incandescent sources

The traditional method of producing light is to heat an object until it is hot enough that the radiation it emits is at visible wavelengths. Examples are fires, candles and incandescent tungsten-filament lamps. For most incandescent sources, the spectral power distribution is similar to one of the blackbody series. The temperature of the closest blackbody is often used as an index of the colour of the source. Typical tungsten-filament lamps are in the range 2500 to 3000 kelvins. The spectral power distribution of a typical tungsten-filament lamp is shown in figure 1.5.

1.3.3 Electric discharge lamps

In electric discharge lamps, radiant energy is produced by passing an electric current through a gas. In most discharge lamps, much of the energy is concentrated in a few narrow bands of wavelength (called spectral lines) in addition to a less intense continuous spectrum. The spectral power distribution of a typical mercury vapour lamp is shown in figure 1.6.

1.3.4 Fluorescent lamps

A fluorescent lamp is a mercury discharge lamp in which the inside of the bulb (usually a long tube) has been coated with phosphors. The ultraviolet radiation from the discharge excites the phosphors to radiate energy of longer wavelengths. The spectral power distribution of a typical fluorescent lamp is shown in figure 1.7. The spectral lines from the mercury discharge are superimposed on the continuous spectrum from the phosphors.

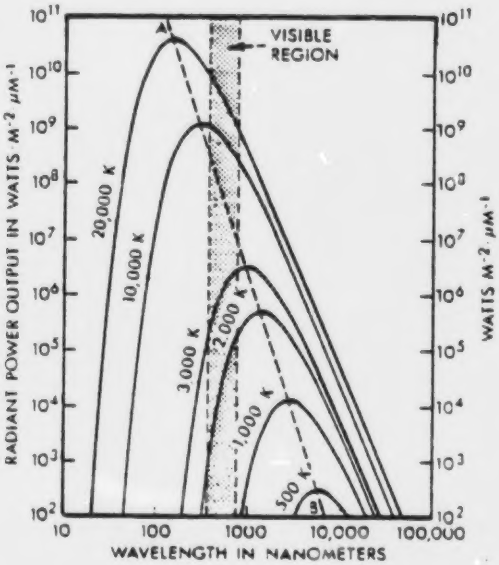

Figure 1.4. Spectral power distributions of blackbodies of temperatures between 500K and 20,000K. Shaded area is region of visible wavelengths (Kaufman, 1984).

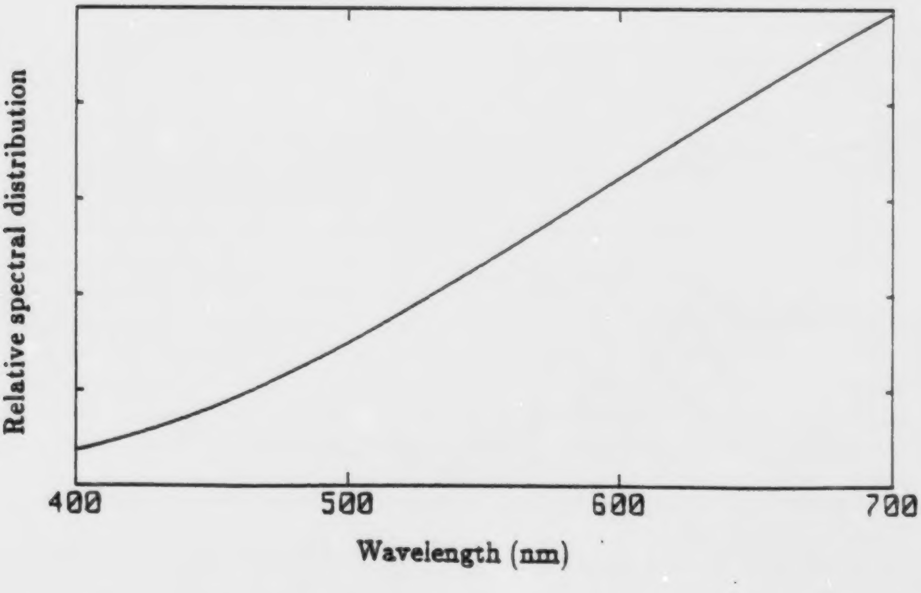

Figure 1.5. Spectral power distribution of a typical tungsten-filament lamp.

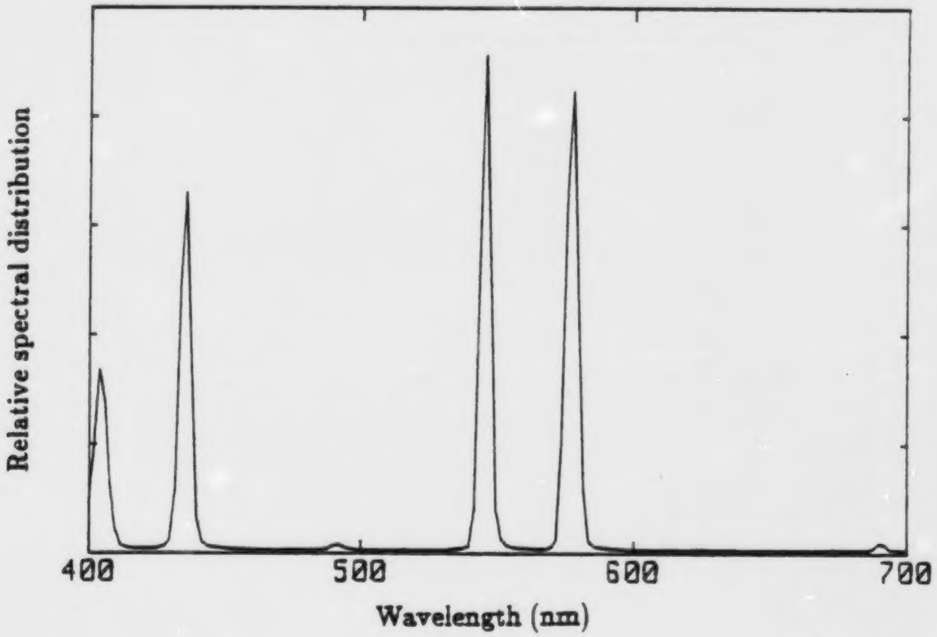

Figure 1.6. Spectral power distribution of a typical mercury vapour lamp.

1.4 Detectors of Light

Detectors of light (optical radiation) are devices that produce an electrical output signal in response to incident light (optical radiation). The electrical signal may be current, voltage, change of resistance, or number of pulses. The wavelength range, size and responsivity (signal per watt of incident radiant power) of various types of detectors are listed in table 1.3.

The relative responsivity of a detector to radiation of different wavelengths is not, in general, the same as that of the human eye. To measure light properly the responsivity must be modified (e.g., with coloured glass filters) or else the detector must be combined with a monochromator (a device to separate optical radiation into its component wavelengths) so that each wavelength (or narrow band of wavelengths) can be measured separately.

1.5 Modulation of Light

1.5.1 Reflection

Reflection is a physical process by which radiant energy incident on a surface or medium is at least partially returned on the incident side of the surface without change of frequency. The reflection can be regular, diffuse, or a mixture of the two.

Regular reflection (sometimes called specular reflection) follows the laws of optical reflection without scattering, as exemplified by mirrors.

Diffuse reflection is reflection in which, on the macroscopic scale, there is no regular reflection.

1.5.2 Transmission

Transmission is a physical process by which radiant energy incident on a surface is at least partially emitted on a side other than the incident side, without change of frequency. As with reflection, transmission can be either regular or diffuse or a mixture of the two.

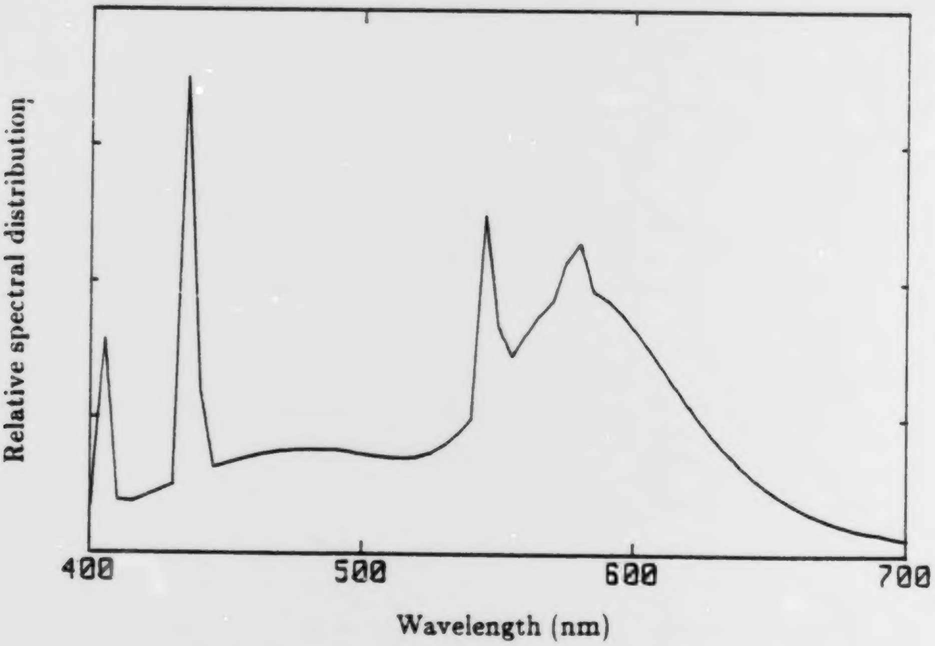

Figure 1.7. Spectral power distribution of a typical fluorescent lamp.

	Wavelength range (nm)	Size (mm^2)	Responsivity (W^{-1})
Thermopile, metal	200–25000	0.8–70	0.06–45 V
Thermopile, thin film	200–25000	0.2–4	13–55 V
Bolometer, thermistor	200–25000	0.01–6	up to 1500 V
Pyroelectric detector	200–25000	1–20	0.3–1.5 μA
Golay detector	200–25000	7–20	2×10^6 V
Vacuum phototube, low V_b	200–1100	25–7000	2–60 mA
Vacuum phototube, high V_b	200–1100	78–18000	10–60 μA/lm
Photomultiplier: S-1	300–1100	80–280	470–2000 A
Photomultiplier: S-4, etc.	180–950	14–12000	$1 \times 10^3 - 3 \times 10^6$ A
Microchannel plate	300–900	120–500	100–10000 A
Gas-filled phototube	200–1100	150–330	125–300 A/lm
VIS photoconductor	350–850	20–500	
NIR photoconductor	350–6000	3–100	$5 \times 10^3 - 5 \times 10^6$ V
Si photodiode, unbiased	250–1100	0.85–800	0.45–0.62 A
Si photodiode, biased	250–1100	0.01–800	0.45–0.62 A
Ge photodiode	400–1900	0.01–80	0.6–2.5 A
Avalanche diode	250–1900	0.007–7	2–85 A

Table 1.3. Wavelength range, size and responsivity of various types of detectors (Budde, 1983).

1.5.3 Fluorescence

Fluorescence is a process by which radiant energy is absorbed and re-emitted as radiant energy of longer wavelengths.

1.6 Instruments

1.6.1 Photometers

A photometer is an instrument for measuring quantities of light. It must have the same relative spectral responsivity as the human eye to light of different wavelengths.

1.6.2 Colorimeters

A colorimeter is an instrument for measuring colour. It must have spectral responsivities that are related to those of the human eye. Two types should be distinguished - those that are designed to measure reflection or transmission, and those that are designed to measure self-luminous sources of light. The first type requires its own internal light source.

1.6.3 Spectroradiometers

A spectroradiometer is an instrument designed to measure radiant energy as a function of wavelength.

1.6.4 Spectrophotometers

A spectrophotometer is an instrument designed to measure reflection or transmission of radiant energy as a function of wavelength.

1.7 References

Budde, W., 1983, *Optical Radiation Measurements, Vol. 4, Physical Detectors of Optical Radiation*, New York: Academic Press.

Kaufman, J.E., 1984, *IES Lighting Handbook - Reference Volume*, New York: Illuminating Engineering Society of North America.

Wyszecki, G, and Stiles, W.S., 1982, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, New York: Wiley.

Notes

Notes

Lecture 2: Wm B. Cowan

BASICS OF HUMAN COLOUR VISION

2.1 Eye and Photoreceptors

This section discusses the optics of the eye, and gives a few properties of the photoreceptors, which are the first stage in the visual process. Wyszecki and Stiles (1982) is a good source for this material, containing many references; Boynton (1979), Boynton (1971), and Hurvich (1981) contain brief summaries. Davson (1972) is a standard treatment of the optics of the eye; Polyak (1941) deals in the same way with the retina, and continues, despite its age to be a standard source.

2.1.1 The eye: optical properties

Imaging

Light enters the eye through the cornea and pupil and proceeds to the retina, as shown in figure 2.1.

Three significant things occur:

1. Light is absorbed by the ocular media: cornea, lens, aqueous and vitreous humours. The absorption is by subtractive colour mixing, so that there is a wavelength-dependent absorption coefficient $A(\lambda)$ and the light reaching the retina has spectral power:

   $$A(\lambda)R(\lambda)\Phi_\lambda,$$

   where Φ_λ is the spectral power of the illuminant and $R(\lambda)$ is the reflectance of the object being viewed.
2. Substantial scattering of light occurs.
3. An image is formed on the retina by the combined action of the cornea and lens. For computer graphics applications it is important to notice that the chromatic performance of the imaging system is poor. In particular, there is substantial chromatic aberration, which means that the eye cannot focus on all wavelengths simultaneously. Suppose the eye is accommodated so that red light is in focus. Then blue light will be defocused so that a point of blue light forms an image 16 minutes in radius. (At 1 meter 16 minutes is 3mm, at 10 feet it is about .35 inches.) This effect makes the blue phosphor in a typical colour monitor, which produces only short wavelength light, quite unreadable except at very short distances.

Visual angle

When we consider the properties of visual stimuli we are usually interested in their location and size on the retina, not in their physical size. Thus we are interested in the angle they subtend, which translates directly into retinal size. The formula which gives visual angle is

$$\begin{aligned}\tan\theta &= \text{`size of object'/`distance to object'} \\ &= \text{`size of retinal image'/`lens-retina distance'},\end{aligned}$$

where θ is the visual angle. The geometry of this calculation is shown in figure 2.2.

The fovea

High acuity vision is possible only in a small portion of the retina, called the fovea. It is about 2° in diameter (35 mm at a distance of 1 metre; 4 in at a distance of 10 ft; an little larger than your thumbnail at arm's length). The fovea is also where colour vision has its highest discrimination.

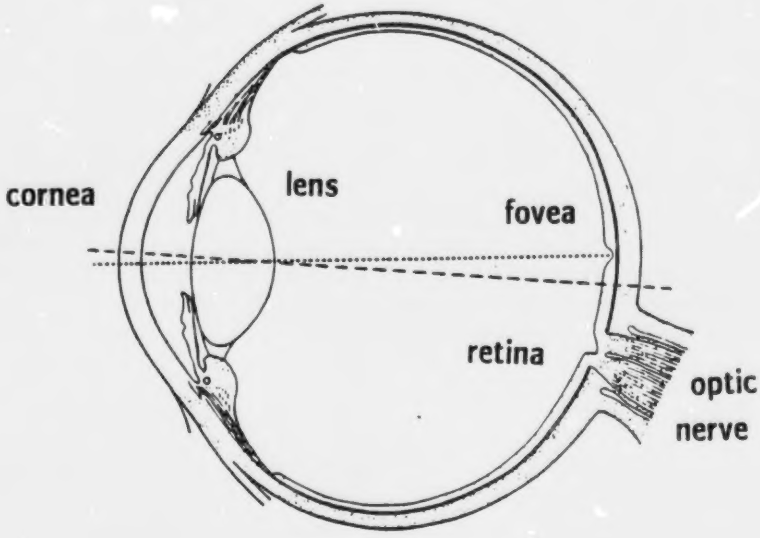
cornea
lens
fovea
retina
optic nerve

Figure 2.1. Schematic illustration of the human eye.

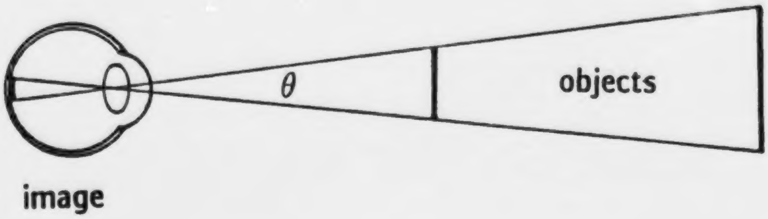
θ
objects
image

Figure 3.2. Illustration of the geometry of visual angle. Two objects at different distances subtend the same visual angle on the retina, provided they are of appropriate size.

The peripheral retina

Outside the fovea vision is capable of less acuity. This effect can be seen in several ways. One is by looking at several printed words some distance away. Fix your gaze on a particular letter, and, without shifting your gaze at all discern your ability to read neighbouring letters. The letter you are fixated on will be sharp, and probably one or two on either side of it, but further away the letters are fuzzy and unreadable. You will find it difficult to maintain your fixation as you do this, as the natural response of the eye-brain system is to move the fovea to any area of the visual field where high acuity vision is required (as it is for reading letters). A second method of getting an idea of the quality of peripheral vision is to think about night vision on a moonless night. After twenty minutes or so it is possible to see general shapes quite well, but impossible to see any fine detail at all. This vision is thought to be similar to that which exists far from the fovea in normal light.

2.1.2 Photoreceptors

Rods

The retina contains two types of photoreceptors: rods and cones. Rods (figure 2.3) play no significant part in normal colour vision. They are unrepresented in the fovea and well-represented in the periphery of the visual field. They are very sensitive at low light levels, but become saturated at normal daylight levels. Their function is to provide high-sensitivity, low-acuity vision, which is used in very low light levels (at night, before the invention of the electric light).

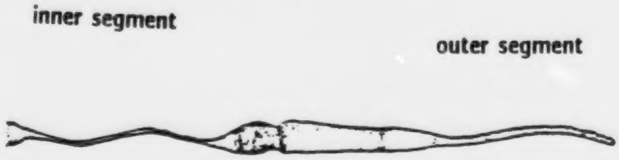

Figure 2.3. Illustration of a human rod. Photons are absorbed in the outer segment; the membrane potential appears on the inner segment.

Cones

Cones (figure 2.4) provide high-detail colour vision at normal light levels.

They exist in three types, which are maximally responsive to short wavelengths (B cones), to medium wavelengths (G cones), and to long wavelengths (R cones). The numerical distribution of the cone types is not even. There are roughly equal numbers of red and green cones, but only about a tenth as many blue cones. This means that blue cones are not as close together as red and green cones. Thus, they are not able to resolve details which are near the limit of the eye's acuity. Such an arrangement makes sense, since the optics of the eye (section 2.1.1) prevent blue light from imaging sharply in any case. Thus, making a blue cone array which could resolve a sharp image which doesn't exist would only be wasteful. Each cone absorbs light in its outer segment. The probability that a photon will be absorbed depends on cone type and on the wavelength of the photon. Call this probability $C_a(\lambda)$, where $a = R, G, B$. Thus, the number of photons absorbed is

$$\int C_a(\lambda)A(\lambda)R(\lambda)\Phi(\lambda)\,d\lambda.$$

The absorption of photons gives rise to an electrical potential on the outer membrane of the cone. The potential depends only on the number of photons absorbed; it does not depend on the wavelength of the

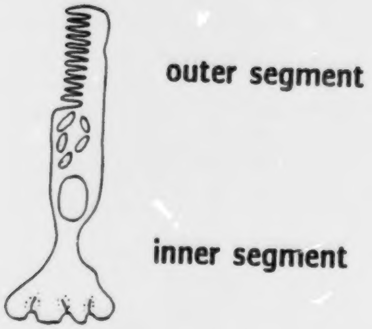

Figure 2.4. Illustration of a human cone. Photons are absorbed in the outer segment; the membrane potential appears on the inner segment.

absorbed photons. The potential does not seem to have any simple functional relation to the number of photons absorbed. As far as is known the cone communicates with the remainder of the nervous system only through this electrical potential. Thus, two lights which produce the same membrane potentials are literally indistinguishable as far as the visual system is concerned. The potential is a one-to-one function of the number of photons absorbed, so that two scenes which have the same number of photons absorbed, in each cone (at each location in the scene) are visually indistinguishable, regardless of the wavelengths of the photons involved.

Several points should be noted:

1. The number of photons absorbed in any cone is a linear function of the spectral distribution entering the eye. Thus applying the same linear operations to two indistinguishable lights leaves them indistinguishable.
2. There are three cone types, so that specifying three numbers (the numbers of photons absorbed, or related values) completely specifies a colour, from the visual point of view. Three numbers do not, on the other hand, suffice to specify an illuminant or a reflectance.
3. For most purposes we are interested in the transfer function between light outside the eye and membrane excitations. In that case cone excitation or related functions ought to include the effect of intraocular absorption so that the products $A(\lambda)C_a(\lambda)$ are given for cone spectral sensitivities. These are the values linear combinations of which are tabulated as colour matching functions (section 4.3.3). They are also the functions shown in figure 2.5.
4. There is even today no general agreement on the correct spectral sensitivities of the three cones (as there is, for example, on the three colour matching functions (section 4.3.3)). One set which is widely accepted is shown in figure 2.5, which uses data determined by Smith and Pokorny (1975).
5. As can be seen in figure 2.5, the red and green cone spectral sensitivities are very similar to one another. For signaling the difference between them, which may be expected to discriminate among colours with signnificant spectral power between 550 nm and 700 nm (i.e. green, yellow, red), the system ought to difference the two cone signals as soon as possible, while the dynamic range of the signals is a maximum.
6. Cones perform a temporal integration of light. The membrane does not respond instantaneously to changes in light, but integrates, with a time constant of about 30 msec (milliseconds).

2.1.3 Things to remember for colorimetry

All of basic colorimetry is based on our good understanding of how photoreception functions in the visual system. From the facts discussed above we can draw several consequences which are very important for colorimetry. They are summarised below.

The principle of univariance

Each photoreceptor produces a signal with only one degree of freedom. The signal depends on the number of photons incident, weighted by wavelength. The consequence is that the output of a photoreceptor, for

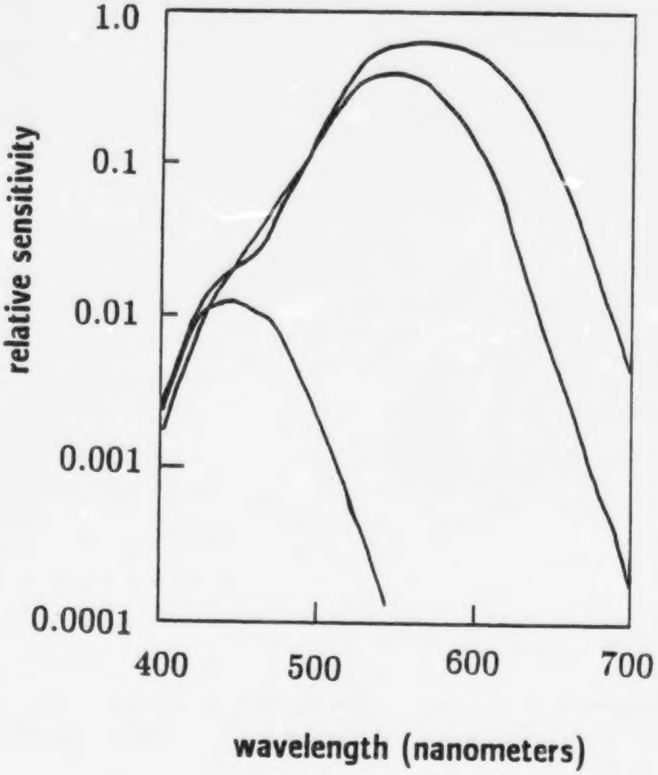

Figure 2.5. Calculated spectral sensitivities of the three human cone types. Note that the ordinate is a logarithmic scale, to encompass the great range of sensitivity displayed by the cones. The spectral sensitivities of the red (R) and green (G) cones are very similar, but that of the blue (B) cone is quite different.

example, does not distinguish between a one photon stimulus at one wavelength and a two photon light at another wavelength to which it is exactly half as sensitive.

Trichromacy

There are three cone types. Thus, when a light stimulates the cone photoreceptors exactly three signals are produced, one by each of the R, G, and B cone types. These three unidimensional signals completely describe the information sent from the photoreceptors to the remainder of the visual system. Thus, we say that colour vision is trichromatic; and that a colour can be described by exactly three numbers, analogous to the signals produced by the three cone types.

Metamerism

Consider a visual system with only one photoreceptor, as our visual system has during rod viewing conditions. Lights of many different spectral compositions produce the same excitation in the single photoreceptor. This effect is called metamerism; two stimuli which are different in spectral composition but produce the same photoreceptor excitation are said to be metamers, or to be metameric one to another. This definition is easily extended to more than one photoreceptor. In the human colour vision system there are three photoreceptors.

Stimuli which are differ in spectral composition but produce the same excitation in each of the three cone types are indistinguishable to the human observer, and are metameric.

Linearity

The number of photons absorbed by any cone type in given by an integral over the spectral distribution of the incoming light. Thus the system is linear with respect to spectral distributions. To be specific, if the spectral distribution of a stimulus is equal to the sum of the spectral distributions of two component stimuli then the number of photons absorbed in response to that stimulus is the sum of the number of photons absorbed in response to each of the two component stimuli. And, if the spectral distribution of a stimulus is a scalar multiple of the spectral distribution of a second stimulus, then the number of photons absorbed in response to that stimulus is the same multiple of the number of photons absorbed in response to the second stimulus. Linearity is important in the derivation of colorimetric laws because it allows the visual response to any stimulus to be represented as the sum of the visual responses to monochromatic stimuli whose strengths correspond exactly to the spectral power function of the stimulus.

2.1.4 Colour defective observers

The discussion of photoreception applies to all 'colour normal' observers. Approximately 8% of males, and less than 1% of females are not colour normal; almost all of these individuals suffer from several colour deficiencies which are well understood. These 'colour defective' observers fall into two categories: anomalous and anopic. Anomalous observers are trichromatic, similar to normal observers in all ways except that one cone, either R or G, almost never B, has an anomalous spectral sensitivity. Anopic observers are not trichromatic but dichromatic, and are missing one cone, either R or G, almost never B. Thus they are unable to discriminate red from yellow from green and are often called 'red-green blind'. More information about colour defective observers can be found in Pokorny et al. (1979), a standard reference on the topic. One fact of general interest is that the inability to make red-green discriminations seems, for anopic observers, to be confined to the fovea (Breton and Cowan, 1981). These individuals have difficulty discriminating small coloured stimuli, such as resistor colour codes, but can distinguish large coloured stimuli, such as the colour of automobiles.

As far as colorimetry is concerned linearity applies to all types of colour defect. But, defective observers separate the set of spectral power distributions into different subsets of metamers than do normal observers. Anopic observers are dichromatic in such a way that any two stimuli that match for a normal observer match for the anopic observer, but that many stimuli that match for the anopic observer differ in colour to the normal observer. Anomalous observers are trichromatic. In general they have different sets of metamers than normal observers, but there is a commonality of two cone types with normals so that some unions of sets of metamers for anomalous observers are the same as unions of sets of metamers for normals.

2.2 Retina, Lateral Geniculate Nucleus and Cortex

Considerable low-level processing of visual information is done by neurons in the retina and in the lateral geniculate nucleus (LGN, a group of relay cells) before the information reaches the visual cortex. Most of what we know about this processing is summed up in the properties of receptive fields, which describe what features of a stimulus excite or inhibit a given neuron. Sections 2.2.1 and 2.2.3 describe a small part of what is known about the receptive fields of these cells, which seem to produce the low level colour coding properties of the visual system.

The processing that takes place in the visual cortex is clearly the most important aspect of the visual mechanism. In recent years it has been the object of a vast amount of research. Section 2.2.4 produces nothing more than a very simple sketch of cortical function, since there is much data on the function of cortical neurons, but few hard results linking that data to psychophysical properties of the human visual system.

The material in this section comes from many diverse sources. There is, as far as I know, no comprehensive survey of this field. Boynton (1971), Boynton (1979), and Hurvich (1981) all give modern, but brief, synopses. Sterling (1983) is a comprehensive survey of retinal processing, but for the cat, not for primates. Similar reviews of research on the cortex, including extensive references to the literature, can be found in Lennie (1980) and Stone et al. (1979).

Receptive fields

The properties of neurons in the visual system are usually determined as follows:

1. A microelectrode is placed in or near the cell while it is functioning normally (as normally as possible!) in a living visual system.
2. The voltage in the electrode, which is assumed to be the voltage on the cell membrane is monitored.
3. Visual stimuli of various kinds: points of light, moving lines, sinusoidal patterns, etc, are placed at different positions in front of the eye.
4. The area in which visual stimuli alter the cell's response is called the neuron's receptive field. The particular patterns which excite the field have no widely used name. 'Trigger features' is perhaps the most common expression.

This procedure is associated with Hubel and Weisel who first used it extensively. There are, however, hundreds of neurophysiologists currently engaged in research of this type. Most low level cells have a centre-surround receptive field organisation. See figure 2.6. There is a small circular area in which a spot of light provokes a strong increase/decrease in spike frequency. This is surrounded by an annulus in which a spot of light decreases/increases the spike frequency. Light in both the centre and surround produces a cumulative response in the form of a signed sum. This receptive field organisation appears in many different versions.

For example, linear or non-linear summation between centre and surround, sensitivity to sustained or transient lights, the same or different cone types contributing to centre and surround, and so on. Three things are important to note:

1. If the centre is excited by light the cell is called on-centre; if it is inhibited the cell is called off-center.
2. Centre-surround organisation is much more sensitive to variations in light than it is to areas that are all the same colour and brightness.
3. Responsiveness to any feature other than position, including colour, because of the centre-surround organisation, is multiplexed with responsiveness to spatial variation. This is thought to be the origin of Mach band and related effects.

2.2.1 The Retina

Bipolar, horizontal, and amacrine cells

These cells comprise the second layer of processing after the cones. See figure 2.7. They are similar to cones in that they give analog responses which vary in time with their input stimulation. They are richly connected to cones and to each other, with feedback circuits included, so that even at this low level complicated processing can take place. Their receptive fields show centre-surround organisation, and when the cell has chromatic sensitivity it is almost always the case that the centre is sensitive to one colour and the surround to another.

Ganglion cells

Ganglion cells represent the final stage of processing that occurs on the retina. See figure 2.7. These cells receive input from the horizontal, bipolar, and amacrine cells, and send signals up the optic nerve to the brain. To do so they have long axons (extensions of the cell) which reach from the eye back into the brain, synapsing onto cells in the lateral geniculate nucleus (LGN). As an adaptation to this long distance signalling, the method used for coding light level changes at the ganglion cells. Cones, and most intermediate cells, use an analogue signal; ganglion cells have fixed amplitude spikes in their membrane potential. The rate of spike production varies as the cell input varies. Most ganglion cells have a resting rate, so that they are able to send negative signals by decreasing spike frequency from the resting rate.

It is interesting to quote some numbers. There are about 100 million rods, all outside the fovea. There are about 5 million cones, of which about 100,000 are in the fovea, where they are most densely packed together. There are about 800,000 ganglion cells, of which about 100,000 relay signals from the fovea. Thus the cone : ganglion ratio is about 1:1 in the fovea, 7:1 outside it. The rod:ganglion ratio is about 150:1. Evidently considerable data reduction occurs right on the retina. Furthermore, concentration on the fovea is already evident at this level.

The receptive fields of ganglion cells are very simple. They are circular, with centre-surround organization. Both on- and off-centre cells are found, and there are cells which respond in both sustained and transient modes. (Sustained cells have a rate of firing which alters when the stimulus is turned on, and which

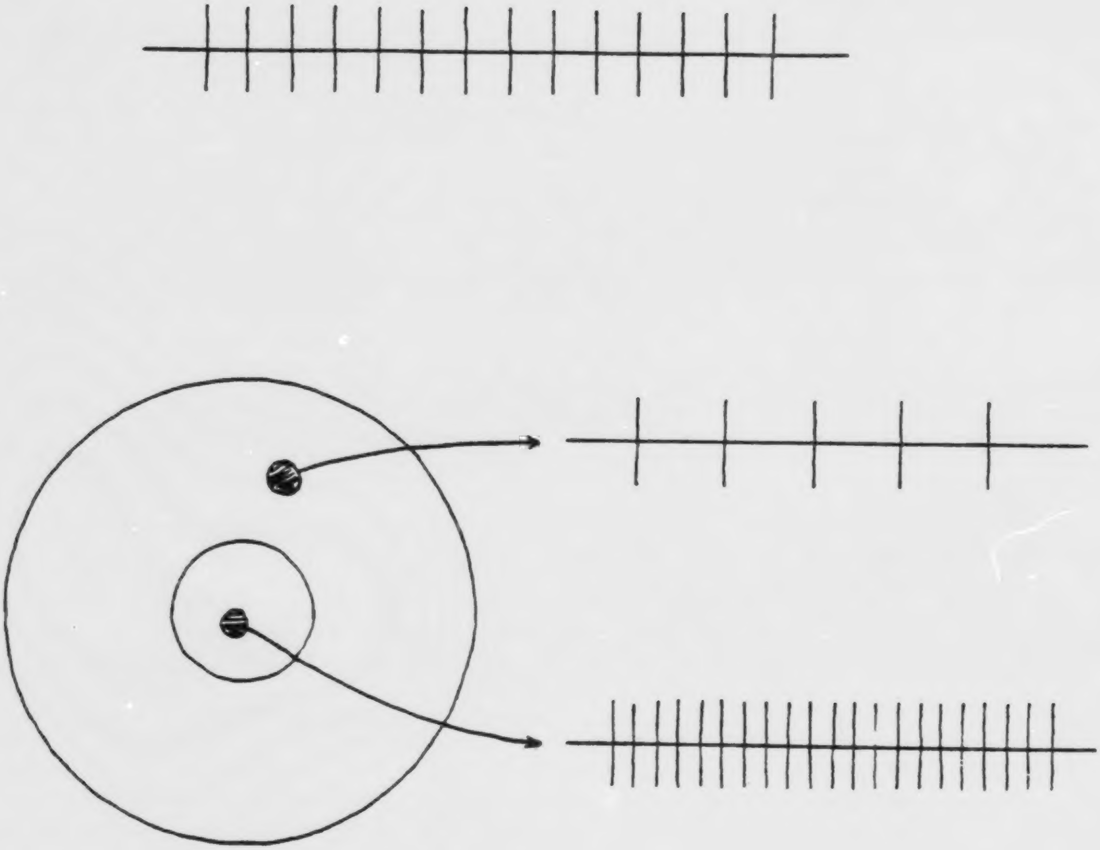

Figure 3.6. Receptive field of a typical retinal ganglion cell. The lines show typical discharge patterns for such a cell: the top is a resting discharge; the centre a discharge when there is a spot of light in the inhibitory surround; the lower a discharge when there a spot of light in the excitatory centre.

remains the same throughout the duration of the stimulus. Transient cells have a burst of firing when the stimulus is turned on and/or off, but soon return to their resting rate when the stimulus stays on steadily.) When ganglion cells are sensitive to colour they are almost always found to have a centre which is sensitive to one cone type, and a surround which is sensitive to another. Cells with centres responding to red cone inputs and surrounds responding to green cone inputs (and the converse) are the most common.

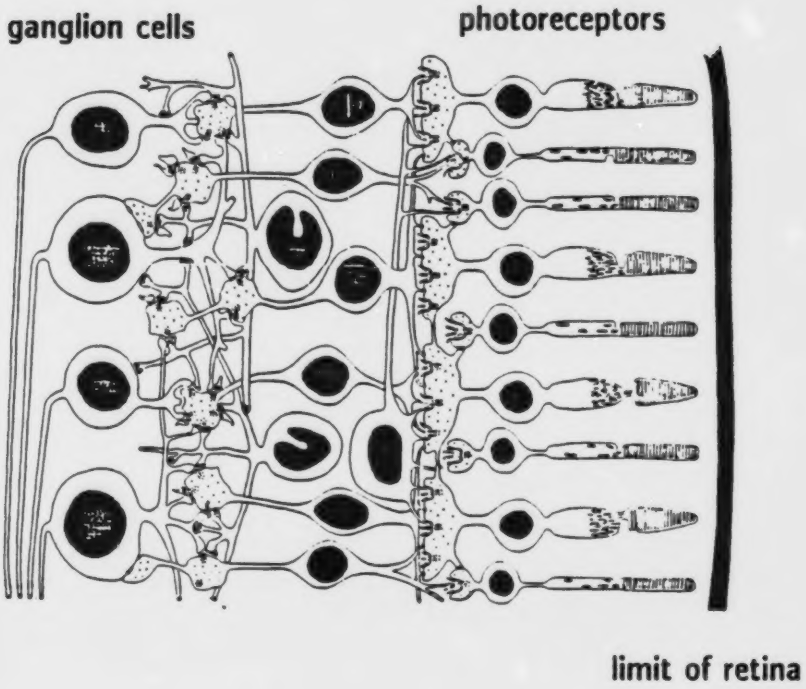

Figure 2.7. Schematic illustration of neural interconnections in the human retina. Information flow is shown by arrows: to the photoreceptors, encoded as light; from the photoreceptors to the retinal ganglion cells, encoded as graded potential; and from the ganglion cells to the brain, encoded as spikes. The intermediate neurons are amacrine, bipolar, and horizontal cells.

2.2.2 The Optic Nerve

The optic nerve is the conduit of visual information from the retina to the brain. Its position is shown in figure 2.8.

It consists of about 800,000 fibres which are the axons of the ganglion cells. Since the activity of 100 million photoreceptors is condensed to less than 1 million fibres, major data reduction has occurred. The optic nerve is, in fact, the major bottleneck for information flow into the visual areas of the brain. Thus, its bandwidth is the bandwidth of visual information flow into the brain. It would be highly desirable to be able to estimate the bandwidth of the optic nerve, but since little is known about how visual information is coded in the nerve it is impossible to do more than speculate.

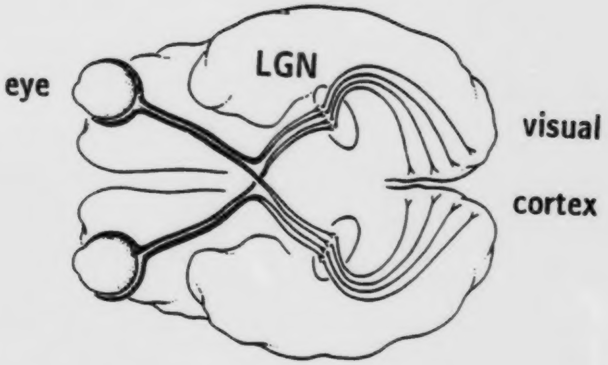

Figure 2.8. Cutaway view of the human brain, showing the eye, optic nerve, lateral geniculate nucleus, and visual cortex.

2.2.3 Lateral Geniculate Nucleus (LGN)

The optic nerves end at the lateral geniculate nuclei, which are shown in figure 2.8. These brain centres are thought to be a relay system, directing signals from the retina to the lower visual centres: superior colliculus, etc. where they control eye movements and the like, and to the visual cortex where analysis of pattern, movement, etc. take place. Receptive field analysis of LGN cells reveals properties which are virtually the same as those of ganglion cells. It is for this reason that the LGN is thought only to relay information, and not to process it.

2.2.4 The Visual Cortex

Overall structure

The location of the visual cortex inside the skull is shown in figures 2.9 and 2.10. It is thought to be the main site of visual processing for complex scenes.

The visual cortex takes input from the LGN, and sends output to almost all other brain centres. Its cells are highly interconnected, with considerable feedback so that it is difficult to say anything simple and concrete about their function and its relation to the cellular structure on one hand, or to visual experience on the other. The visual cortex is generally divided up into several areas, named Brodmann's areas 17, 18, and 19. These areas are thought to be specialized, and to have different information processing functions, although there is no consensus on exactly what the functions are.

Area 17

Area 17 receives information directly from the LGN, and relays it into areas 18 and 19, as well as to other parts of the brain. It is the most intensively studied part of the cortex. The most interesting concept in its structural organization is that of the retinotopic map. This refers to the fact that neurons in area 17 are laid

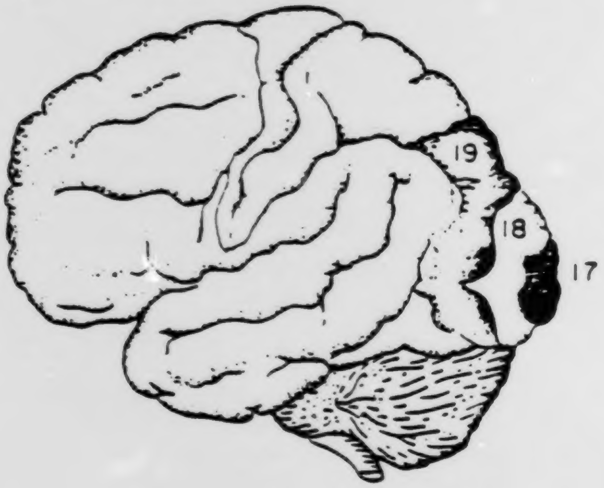

Figure 2.9. Cutaway view of the human brain, showing the position of the visual cortex. The brain is shown from the side, with the rear to the right. The areas labelled 17, 18, and 19 are the corresponding parts of the visual cortex.

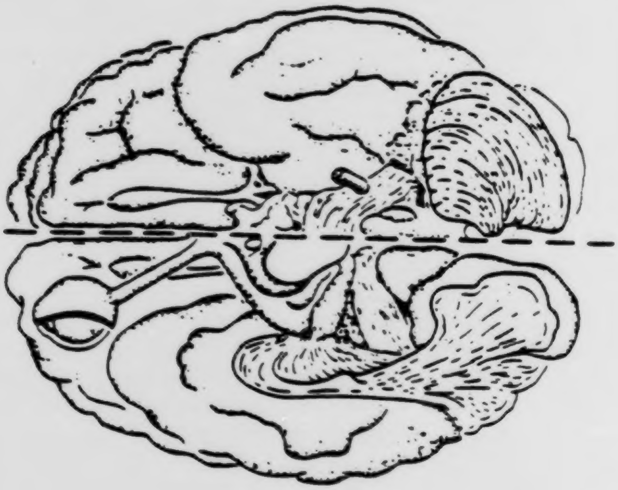

Figure 2.10. Cutaway top view of the human brain, on the same scale as figure 2.9, and on the same scale and orientation as figure 2.8. See the labels of figure 2.8 for the different parts of the visual system which are shown.

out so that there is a topological correspondence between positions of a cell in area 17 and the part of the visual field in which its receptive field is found. The left half of the visual field is represented on the right side of the cortex, and the right half of the visual field is represented on the left side of the cortex.

Areas 18 and 19

Areas 18 and 19 receive input from area 17 as well as from the LGN directly. They are thought to be involved in higher level processing than is area 17, though there is some controversy on this point. Each area contains several retinotopic maps, each containing a complete representation of half the visual field, with the right half on the left side of the cortex, and vice versa. It is thought that each different map is associated with the extraction of a particular type of visual information. By this time we have moved into the realm of sheer speculation, so it is time to say no more.

Cell numbers

One important fact is the number of cells. There are estimated to be 320 million cells and 740 billion synapses in area 17 alone (O'Kusky and Colonnier, 1982). (This estimate is for macaque monkey, which has a visual system very similar to the human visual system. The numbers are probably a significant underestimate when applied to humans.) There is obviously an enormous processing capacity in the cortex, especially when it is considered that area 17 is considerably the smallest of the cortical areas.

Receptive fields

As might be expected the receptive fields found in area 17 are quite similar to those found in the retina and in the LGN. Thus, we find centre-surround organised receptive fields, both on- and off-centre, both transient and sustained. Areas 18 and 19 have more complex receptive fields, showing sensitivity to orientation, colour, motion, and so on. Despite much effort there has been no consensus on which of the several competing classification schemes represents best the complexity of cortical processing. One particular problem is that much work has been done on the cat and lower animals, only a little on primates, and it has turned out that while the general principles of visual system organisation are quite similar throughout all mammals, only the primates have visual systems which are organised in the same manner that human visual systems are.

Colour

Cells sensitive to colour are found throughout area 17, and also in areas 18 and 19. They are, in general terms, organized like cells in the LGN and the retina. There is no good theory of how colour information is processed by them. Such a theory must be thought to be some considerable distance in the future.

One intriguing possibility which has been mentioned is worthy of notice. It stems from the fact that there is a retinotopic map, called V4, in which several studies have found very large proportions of colour sensitive cells, much higher proportions than are found in other parts of the cortex. Here, it is suggested, is the place where colour is added to the visible world. This speculation is reinforced by the existence of several brain damaged patients who have lesions localised to V4. These people report seeing a world which has no colour, describing the difference before and after their injuries as being the difference between a colour and a black and white movie. The difficulty of work in this area of research means, unfortunately, that these ideas are likely to remain no more than an intriguing speculation for many years.

2.3 Processing of Colour Signals

This section describes what is currently believed about the nature of how colours are coded by the human visual system. This knowledge is primarily psychophysical, derived from experiments in which the human response to light is studied. Boynton (1971) and Boynton (1979) are both good introductions to the material in this section. Hurvich (1981) is probably the best reference. It is particularly strong in the area of opponent colours, of which its author is one of the leading protagonists.

2.3.1 Luminance: flicker photometry

The facts about colour matching show quite clearly that colour experience has three degrees of freedom. It is natural to suppose that they consist of three separate channels within the visual system. Given the possibility of three separate channels it is natural to ask what aspects of colour experience are produced by each one, with the idea that these channels will represent a real perceptual basis for colour space. One technique which has been used to discover individual channels is to suppose that, if the channels are indeed separate, they may have different response properties (for example, to spatial or temporal stimulus variation). Such a strategy, using temporal response, has been successful in isolating what is called the luminance channel.

Flicker photometry

It turns out that there is a channel distinguished by its temporal response; its characteristics are revealed by flicker photometry. This is a stimulus arrangement where two lights are alternated at a slow rate. This stimulus is shown schematically in figure 2.11.

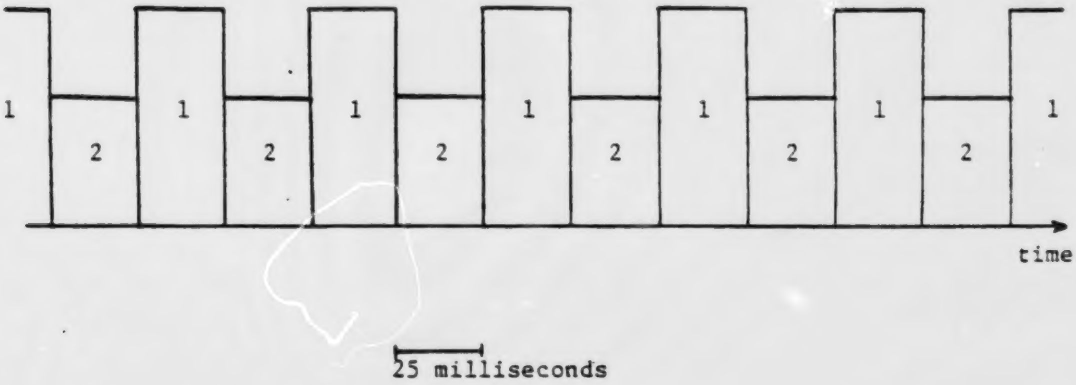

Figure 2.11. Stimulus for a flicker photometric measurement which alternates colours 1 and 2. The level of one colour (e.g. 2) is varied, its hue and saturation remaining constant, until the sensation of flicker is a minimum.

Suppose that the lights are red and green. If the rate is very slow then an observer sees red, then green, then red, and so on, where each colour is present for a visually discrete time period. On the other hand, if the rate is very fast, the observer sees red and green additively mixed into a steady yellow colour, and has no sensation at all that alternation is taking place. Between these two extremes there is an intermediate rate, somewhere around 20 Hz (cycles per second; which means that there is 25 milliseconds of red, followed by the same period of green, and so on), where a qualitatively different sensation is perceived. The light is of somewhat indeterminate colour, and there is a strong sensation of flickering. The question is: by what mechanism, and on what channel or channels is the flickering perceived?

The question is answered in a manner very similar to the technique used to prove trichromacy: if the observer can remove the flicker by varying a single degree of freedom, then the flicker must be perceived from a single channel; if he needs two, then two channels must contribute; and if he needs three, then any aspect of colour difference can contribute to flicker, and all three channels are implicated. Let us suppose that the red light is to be varied. First, it is quite clear that flicker can be removed if the red light is varied in three degrees of freedom, because with all degrees of colour variable the red can be changed to a green of exactly the same hue, saturation, and brightness as the green light. No flicker will be perceived because there is a steady green colour. Thus, the question of interest is: can flicker be removed when only one or two degrees of freedom are variable?

When we do the experiment we discover that only one degree of freedom is necessary to remove the sensation of flicker. That is, if we allow the observer to vary only the intensity of the red light, while its hue and saturation are held constant, he can find a unique intensity where the sensation of flicker vanishes almost completely. Thus, only one of the three colour channels is sensitive to flicker at the given rate. It is called the luminance, or achromatic, channel, and its properties are investigated using the technique of flicker photometry.

The luminance channel

The properties of the luminance channel can be enumerated as follows:

1. It is linear with photon number. Thus it obeys Grassmann's second law. i.e. The luminance of a linear additive mixture of two lights is the same linear combination of the luminances. This suggests either that luminance is a pigment property, or that it is formulated very close to the pigment stage of the visual system (i.e. very close to the photoreceptors, before non-linear transformations have taken place.) For reasons which we will not go into here the second interpretation is generally preferred.
2. It has a relatively broad spectral response, being relatively sensitive to photons of all wavelengths. Hence it is called achromatic.
3. It is a linear combination of the colour-matching functions.
4. Its spectral sensitivity is called $V(\lambda)$, which is widely tabulated, and was standardized by the CIE in 1924 (though not at the exact value currently believed to be correct).
5. The contribution of a light to this channel, its luminance, is generally called Y, and is calculated from

   $$Y = \int V(\lambda)\Phi_\lambda \, d\lambda$$

   where Φ_λ is the spectral power distribution of the light. (See lecture 4 for constants to get the right units in the definition of Y.)

2.3.2 Colour naming

Flicker photometry appears to isolate one degree of freedom (or channel) in the three-dimensional colour response. It has broad band properties, and seems to be associated with determination of brightness, not of hue or of saturation. Now we would like to consider how the visual system might encode the hue and saturation. A useful method of considering this problem looks at the phenomenon of colour naming.

Four colour names

With colour names we will follow the course plotted in colour matching once again. Give an observer a wide variety of colour samples, and the task of naming them. To begin with, allow him one colour name e.g. red. You will find that no matter what name he is given he will have many samples that cannot be named. Then, allow him two colour names, e.g. red and yellow, and also allow him to use combination of the two, yellowish-red, reddish-yellow. He will be able to name more of the samples, but many will be unnamable. Allowing three names improves things further, but things are still unsatisfactory. But, when four are allowed, provided that the four are red, yellow, green, and blue, you will find him able to name virtually all the samples. The details of how to control such an experiment are too contentious to mention. The general finding, however, that even when the observer is reluctant to name a colour he will, when pressed, do so, and in a consistent way from day to day.

The colour circle

The four names seem to suggest that there are four degrees of freedom when describing hue and saturation, when we were looking for only two. Closer examination of the names, however, demonstrates what is happening. With four colours, six combination names, e.g. red-yellow, are possible. Only four occur in the naming; they are:

red-yellow,

red-blue,
green-yellow,
green-blue.

The possible combinations:

green-red,
yellow-blue,

never occur. This circumstance suggests that the four names actually represent two polarized axes. One is red in one direction, green in the other; the second is yellow in one direction, and blue in the other. With achromatic colours in the centre, where both red/green and blue/yellow are simultaneously zero, this coordinate system suggests the familiar colour circle or hue circuit, figure 2.12. Around the circle are segments corresponding to each of the occurring combination names, but no position exists for those combinations that fail to occur.

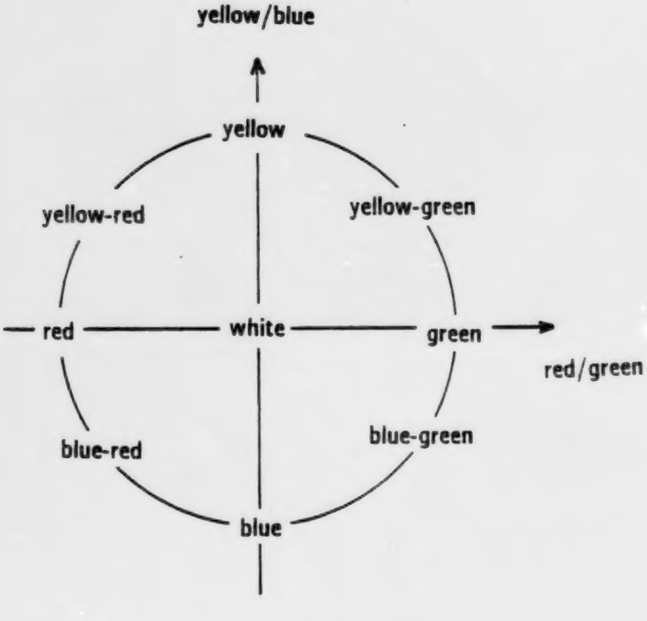
yellow/blue
yellow
yellow-red
yellow-green
red
white
green
red/green
blue-red
blue-green
blue

Figure 2.12. Arrangement of colour names around the hue circuit.

2.3.3 Unique Hues

To reinforce the points made in the previous section, let us consider the rather remarkable phenomenon known as the unique, or unitary hues.

Unique yellow

Consider a range of colours in the neighbourhood of yellow. They can be put on a continuum from the most reddish yellow to the most greenish yellow. Somewhere in the middle of the range is a colour which is neither reddish nor greenish, but just yellow. That colour is called unique yellow. If we ask many observers to select a yellow colour that is 'neither reddish nor greenish' we find two rather surprising facts:

1. The colour chosen is surprisingly stable when we get the same observer to do the task over many days. (For this stability to occur we must take care that the observer does the selection under consistent adaptation conditions. He should not, for example, select the colour immediately after looking at red for ten minutes).
2. There is surprisingly little variation from observer to observer. The standard deviation of the variation is the same general size as the minimal discriminable hue shift at this wavelength.

These two facts lead us to suspect that unique yellow is not a cultural or developmental phenomenon, but rather a consequence of the basic wiring of the visual system.

Unique green

As well as unique yellow, there also exist unique red, green, and blue, which have properties similar to unique yellow. Unique green is unusually interesting, and provides further evidence that the unique hues are low-level phenomena. There is considerable evidence that unique green judgments fall bimodally in the population, with the separation between the modes larger than the standard deviation of either. The distribution is shown in figure 2.13.

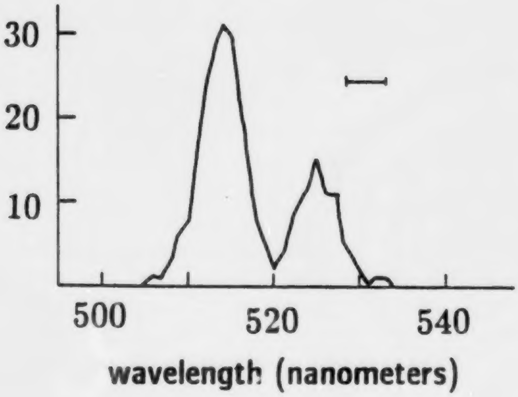

Figure 2.13. Histogram of wavelengths chosen for unique green by a large population of observers. The smallest wavelength difference that can be discriminated in this region of the spectrum is shown by the horizontal bar.

Furthermore, which mode a person falls into is strongly correlated with the following properties of his visual system:

1. the colorimetric coordinates of white,
2. the additivity of spectral lights, and

3. the recovery of sensitivity following adaptation to an intense light,

all of which are bimodally distributed through the population (Richards, 1967). The strong correlations of these seemingly basic judgments once more suggests that unique hues are connected with the fundamental wiring of the nervous system.

The colour circle

The unique hues have a simple interpretation in terms of the colour circle and red/green and blue/yellow colour channels. Simply put, unique yellow and blue occur when the red/green channel is at its null point; unique red and green occur when the blue/yellow channel is at its null point; and achromatic colours appear when both channels are null. Thus, the low level nature of the unique hues implies that the red/green and blue/yellow colour channels represent a low level organisation of colour signals.

2.3.4 The opponent colours model

We can put together all the phenomena discussed so far and get the model that is generally believed by most psychophysicists. It is called the opponent colours model, and was first put forward in its present form by Hurvich and Jameson in the 1950's. Its main components are:

1. Light is absorbed by three types of cones: a long wavelength R cone, a medium wavelength G cone, and a short wavelength B cone.
2. Signals from the R and G cones are summed to form an achromatic signal. 'Summed' in this context, as with 'differenced' below, should not be understood to be a precise arithmetic operation. It means only that the signals are put together cumulatively so that positive R cone increments and positive G cone increments both affect the achromatic channel positively.
3. Signals from the R and G cones are differenced to form the red/green channel.
4. Signals from the achromatic channel and from the B cones are differenced to form the blue/yellow channel.
5. The structure just described is repeated at every point in the visual field.
6. The structure described assumes that there are 'layers' of processing which can be assumed to occur in different zones of the visual mechanism. For this reason it is one of a family of 'zone' theories.

This model, summed up by figure 2.14, is currently accepted as a good first cut at a picture of how colour encoding works. It is, of course, supported by many psychophysical experiments besides the ones mentioned here. In addition, there is some physiological evidence that cells having opponent colours response properties do exist in the LGN.

Much current research in the area of colour psychophysics is involved with working out the details of the opponent colours model. For example,:

1. The exact mathematical form of the transformations from cone signals to the colour channels,
2. The relationship between colour encoding and other phenomena, such as simultaneous contrast, adaptation, and so on,
3. The identification of the exact physiological components that perform the opponent colours transformation,
4. And so on.

All the same, there are phenomena which suggest that the opponent colours model is not the whole picture. These include:

1. the observation that perceived brightness is not the same as luminance,
2. the close interaction of spatial and colour degrees of freedom in the visual system (suggesting that zone theories are not tenable),
3. the natural separation of colour into hue and saturation,
4. cognitive phenomena, and
5. the nature of brown.

It is probably true to say that vision scientists will eventually settle on a multi-representation model, where different colour representations are used for different tasks. It is also probably true that the opponent colours model will continue to be the best candidate for the front end of the multi-representation model.

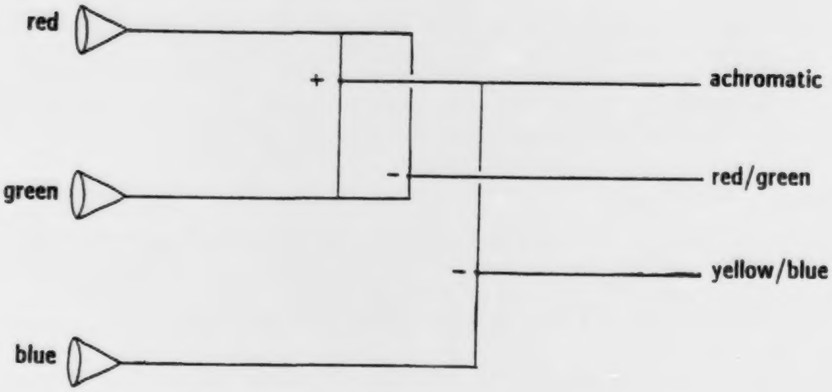

Figure 2.14. Schematic diagram of neural connections postulated by the opponent colours model. Red and green cone signals sum to make the achromatic channel, difference to make the red/green channel. Achromatic and blue cone signals difference to make the yellow/blue channel.

2.3.5 Psychophysics of the cortex

We have associated the psychophysics of colour matching with properties of cones. It is reasonable to associate the psychophysics of opponent colours with the remainder of the retina and with the LGN. Is there any psychophysics which can be associated with the cortex? It is quite possible that the spatio-temporal properties of colour channels, not discussed in this lecture, are as much of cortical origin as they are of retinal and geniculate origin. Other than that the cortex is a sort of grab bag. We know little enough about it that we can attribute virtually any psychophysical phenomenon which cannot be explained elsewhere to it. Thus, the list which follows is not much more than a list of things we don't understand very well:

1. Contrast effects, colour induction, and colour constancy.
2. Colours like brown, navy blue, olive green, which can only be seen in complex scenes.
3. Brightness, which is not, as it turns out, the same as luminance.
4. Adaptation effects, whereby the eye takes illumination into account when trying to discover the properties of surfaces it is examining.
5. Colour spreading effects, illusory colours.

This list is just a sample. There are literally hundreds of colour facts which must be ultimately explained by recourse to the properties of cortical circuitry.

2.4 References

Boynton, R. M., 1979, *Human Color Vision*, New York: Holt, Rinehart, and Winston.

Boynton, R. M., 1971, 'Color Vision', in *Experimental Psychology*, ed. Kling, J. W. and Riggs, L. A., New York: Holt, Rinehart and Winston.

Breton, M. E. and Cowan, W. B., 1981, 'Deuteranomalous color matching in the deuteranopic eye', *Journal of the Optical Society of America*, **71**: 1220–1223.

Davson, H., 1972, *The Physiology of the Eye*, New York: Academic.

Hurvich, L. M., 1981, *Color Vision*, Sunderland: Sinauner Associates.

Lennie, P., 1980, 'Parallel Visual Pathways: a Review', *Vision Research*, **7**: 561–594.

O'Kusky, J. and Colonnier, M., 1982, 'A Laminar Analysis of the Number of Neurons, Glia, and Synapses in the Visual Cortex (Area 17) of Adult Macaque Monkeys', *The Journal of Comparative Neurology*, **210**: 278–290.

Pokorny, J., Smith, V. C., Verriest, G. and Pinckers, A. J. L. G., 1979, *Congenital and Acquired Color Vision Defects*, New York: Grune and Stratton.

Polyak, S. L., 1941, *The Retina*, Chicago: The University of Chicago Press.

Richards, W., 1967, 'Differences among Color Normals: Classes I and II', *Journal of the Optical Society of America*, **57**: 1047–1055.

Smith, V. C. and Pokorny, J., 1975, 'Spectral Sensitivity of the Foveal Cone Photopigments between 400 and 500 nanometers', *Vision Research*, **15**: 161–171.

Sterling, P., 1983, 'Microcircuitry of the Cat Retina', *Annual Reviews of Neuroscience*, **6**: 149–185.

Stone, J., Dreher, B. and Leventhal, A., 1979, 'Hierarchical and Parallel Mechanisms in the Organization of the Visual Cortex', *Brain Research Reviews*, **1**: 345–394.

Wyszecki, G. and Stiles, W. S., 1982, *Color Science*, second edition, New York: Wiley.

Notes

Lecture 3: Wm B. Cowan

BASICS OF COLORIMETRY

What is colorimetry? One thing that is quite clear is that the perceptual world is divided into regions of colour, some of which are the same in colour and some of which are different. What physical characteristics of the light emitted by an object causes it to look the same or different in colour from another object? Is it possible to design an instrument which can examine a surface and determine whether a human observer will conclude that it is the same or different in colour from another surface? Colorimetry attempts to provide answers to such questions, and is successful at doing so within a rather narrow range of visual experience. The sections which follow describe the theoretical foundation by which the most simple of colorimetric questions can be answered.

The experiments which underly the functions used in the practice of colorimetry are ones which examine the properties of colour mixing and colour matching. How to do them was first elaborated in the late nineteenth century; and the actual experiments on which our present standards are based were done in the first half of the twentieth century. This lecture summarizes those experiments, the concepts on which they are based, and the general properties of colour mixing and matching which they discovered. Wyszecki and Stiles (1982) is the standard source for the material in this section. Of course, as soon as a framework exists to respond to a simple question (Are two coloured stimuli the same in colour when seen in identical visual surroundings?), we want to ask more complicated questions (What is the smallest change in colour that is visible? How does a colour change when the visual surroundings are changed?). The standard methods developed here are less successful at providing answers to the more complicated questions. Thus, the final section of this lecture will outline some of the limits of the formalism of basic colorimetry.

3.1 Colour Matching

When do two colour stimuli match one another? The answer to this question is bound up with the concept of metamerism, first mentioned in section 2.1.3 of lecture 2.

Metamerism

Suppose that the eye views two coloured areas or lights. If the two emit light of the same spectral distribution, then exactly the same photons enter the eye from each one, all physical, chemical, biological, and neural events which occur inside the eye and visual system must be exactly the same (given identical states of adaptation for the visual system, which can be assured by identical histories of visual stimulation) and the same colour will be seen from each. What if the two spectral distributions are not the same? In this case the colours may be seen to be either the same or different. (Note that when we say 'same' here we are not considering the finite precision of the visual system; the colours we indicate as being the same would be the same even in a visual system of infinite precision.) In this case where spectral distributions are different we call them 'metameric' if they are the same colour. Thus, two different spectral distributions are perceived to be the same colour (under identical adaptation conditions and in identical visual environments) they are called 'metamers', and the match between the colours is called a 'metameric match'. By contrast, when two colours match because their spectral distributions are the same the match is called a 'physical match'.

Metamerism arises because the mapping from spectral distribution to colour is not one-to-one. Many spectral distributions correspond to the same colour. In other words, the transformation

$$\Phi_\lambda \Rightarrow C \tag{3.1}$$

where Φ_λ is the spectral distribution of a stimulus and $\mathbf{C}$ is the colour of the same stimulus is many to one. A typical example of metamerism is found in colour television. Many colours in the environment can be matched exactly on a colour television monitor. In the environment the colour is created by a spectral power distribution in which any wavelength can vary independently of the others. On the television screen it must be a sum of the spectral power distributions of only three phosphors. Almost all such matches are therefore metameric.

Notation

To indicate that two spectral power distributions are the same we write

$$\Phi_\lambda = \Phi'_\lambda \tag{3.2}$$

indicating that they are the same for every value of the wavelength λ. To indicate that two colours are perceived to be the same, or that they match we will write

$$\mathbf{C} \sim \mathbf{C}' \tag{3.3}$$

where the symbol $\sim$ is understood to mean 'matches'. Note that because of metamerism equality (3.2) implies match (3.3), but not the reverse.

3.2 Colour Mixing

There are two methods of mixing colours: additive and subtractive. Most people are more familiar with the subtractive method, which is the method by which colours mix when paints are mixed. The properties of the additive method, which occurs in the colour generation process used by a colour television monitor, are much more simple, so that additive mixture has been made the basis of colorimetry. In this section we will discuss in detail the properties of additive mixture, which will be the basis of our further investigation of colorimetry. At the end is a much more brief discussion of the more complicated subject of subtractive mixture.

3.2.1 Additive colour mixing

It is possible to use a beam-splitting plate to make light from two sources combine so that they excite exactly the same point on the retina. This is illustrated in figure 3.1.

In such a case the colours of the two sources combine into a new colour. Such a colour combination is called additive mixture, because photons from the two sources add together to give the combined spectral power distribution. That is,

$$\Phi_\lambda = \Phi_{1\lambda} + \Phi_{2\lambda}, \tag{3.4}$$

where $\Phi_{1\lambda}$ and $\Phi_{2\lambda}$ are the two spectral power distributions of the two components and Φ_λ is the spectral power distribution of the combination. Additive mixture has simple and useful properties which we will examine in the next few sections. First, let us look at some of the places where it occurs in everyday visual experience.

Additive mixture occurs frequently because of the limited spatial and temporal resolving power of the human eye. That is, the eye averages light received from a small area of the visual field (the order of 1 to 10 minutes in extent). Light from anywhere in that area is seen as if it were coming from the same place, and all the photons are combined additively. This is taken advantage of in the design of:

1. Colour television monitors - the individual red, green and blue phosphor dots are so close together that the eye sees the additive mixture of all three, which forms a complete range of colours.
2. Offset printing - different colours are put onto several screens which are printed in a very fine pattern on the paper. Elements from different screens are combined, because of spatial proximity, into the impression of a range of colours.
3. Painting - the pointillist school of painting (e.g. Seurat) used many small dots to form the picture. The dots combine additively into a full range of colour.
4. Fabric design - Thread of many colours is woven together into a complicated pattern. From close up the individual colours can be seen, but from farther away the colours combine into a single overall colour.

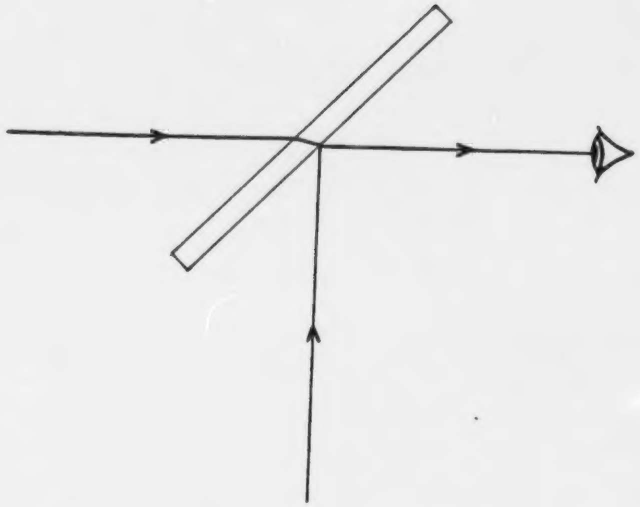

Figure 3.1. Additive mixture of two photon sources. A beamsplitter, which reflects light which comes from one side, while transmitting light which comes from the other, makes photons coming from two different sources (directions) excited the same point on the retina. Thus, light from the two sources is additively mixed. When viewing such an arrangement the observer would see two images superimposed on one another (a double exposure).

There are many other examples of additive mixture through spatial contiguity, including the computer graphics technique of spatial dithering. Because the eye also has finite temporal resolution it is also possible to make additive mixtures by rapid alteration in time of different colours (the order of 30 milliseconds for a full cycle is appropriate, see section 4.2 for more details). This is not very often seen in practice, but any television screen is an example, since any given pixel is almost always dark but is illuminated briefly by a short burst of light. That burst is combined with the darkness by additive mixture to create the colour we see.

The actual colour mixing properties of additive mixture are best discovered by experimentation, which is easy to do using a computer graphics system. They are the properties familiar from stage lighting: red plus green makes yellow, blue plus yellow makes white, blue plus red makes violet, and so on.

Notation

Equation (3.4) gives the notation used to specify spectral power distributions under the additive mixture of two stimuli. To specify the additive mixture of colours under additive mixture we will use the symbol $\oplus$. Thus, the colour equivalent of (3.4) is

$$\mathrm{C}_1 \oplus \mathrm{C}_2 \sim \mathrm{C}. \tag{3.5}$$

3.2.2 Subtractive colour mixing

Subtractive colour mixing is a much more complicated phenomenon than additive mixing, so that it is not possible to give a detailed account of how it occurs. The best we can do in a short space is to describe how subtractive mixing occurs with filters, and give a short discussion of how the treatment might be extended to paints.

Recall that light can become coloured as a result of passing through a coloured medium (which we call a filter). After light has been coloured by passing through a filter, the spectral power distribution of the light is

$$\Phi'_\lambda = T_1(\lambda)\Phi_\lambda, \tag{3.6}$$

where $T_1(\lambda)$ is the spectral transmission function of the filter. If we pass the resultant light through a second filter the resulting spectral power distribution is

$$\Phi''_\lambda = T_2(\lambda)T_1(\lambda)\Phi_\lambda, \tag{3.7}$$

where $T_2(\lambda)$ is the spectral transmission of the second filter. It is possible to extend this multiplicative treatment of light absorption, deriving the concept of optical density, but that goes beyond our present purpose. We wish only to note two points:

1. Spectral power distributions produced like this can be subjected to the sort of analysis given below (sections 3.3 to 3.5), but no simple results will be derived. This occurs because subtractive colour mixing is essentially multiplicative while photoreception, on which colorimetry is based, is additive.
2. For the same reason there are no equivalents to Grassmann's Laws (section 3.3.3) for subtractive colour mixing.

The question of how pigments in paints combine is much less simple than filters. It suffices to note that a theory of the spectral distribution which results when two paints are mixed depends on having a detailed theory of how pigment particles are suspended in the medium, and of how multiple scattering of light off pigment particles within the medium occurs. These cannot be simply addressed. Furthermore, even if there were a complete theory of the chemistry and physics of paint it would have relation to colorimetry as described below, and no equivalent of Grassmann's Laws.

Because subtractive colour mixing depends on the whole spectral power distribution, it is not possible to predict the results of subtractive colour mixing without knowing the whole spectral power distribution of the colours being mixed. And since human colour response depends only on three degrees of freedom (because of trichromacy) it is not possible to predict the results of subtractive colour mixing on the basis of the colour identification procedures described in section 3.4. (In this subtractive mixing differs from additive mixing, the results of which can be predicted on the basis of the human colour response to the colours being mixed. The preservation of metamerism under additive mixture is what allows this. Metamerism is not preserved under subtractive mixture.) We would like to give a few examples of the results of subtractive colour mixture, to compare to the results of additive mixture. Adding red and yellow gives red, or orange, or brown/black. Adding red and green usually gives brown or grey/black. Adding red and blue usually gives purple or grey/black. Adding yellow and green usually gives green. And adding yellow and blue usually gives green. Unfortunately none of these rules has universal application, as anyone who has worked with paint knows very well.

3.2.3 Subtraction in additive colour mixing

It is important to avoid confusing subtractive colour mixing, described immediately above, with subtraction, $\ominus$, in the context of additive colour mixing, described in section 3.3.2 below.

3.3 Metamerism and Additive Mixture

Significant regularities occur when metamerism and additive mixture are taken together. Additive mixture sums photons, which amounts to adding together spectral power distributions. Thus, we can express the effect on the spectral power distributions of additively mixing two stimuli as in equation (3.4)

$$\Phi_{3\lambda} = \Phi_{1\lambda} + \Phi_{2\lambda}. \tag{3.8}$$

Now, $\Phi_{1\lambda}$ belongs to metamerism class $\mathbf{C}_1$, which means that it is perceived as the colour $\mathbf{C}_1$. Similarly, $\Phi_{2\lambda}$ is colour $\mathbf{C}_2$ and $\Phi_{3\lambda}$ is colour $\mathbf{C}_3$. We can write the fact that the additive mixture of $\mathbf{C}_1$ and $\mathbf{C}_2$ matches $\mathbf{C}_3$ by:

$$\mathbf{C}_3 \sim \mathbf{C}_1 \oplus \mathbf{C}_2, \tag{3.9}$$

where $\sim$ means matches, and $\oplus$ means additive mixture. There is a pitfall lurking in this notation. It is possible that if we chose $\Phi'_{1\lambda}$ metameric to $\Phi_{1\lambda}$ and $\Phi'_{2\lambda}$ metameric to $\Phi_{2\lambda}$, that $\Phi'_{3\lambda}$, the additive mixture of $\Phi'_{1\lambda}$ and $\Phi'_{2\lambda}$, would not be metameric to $\Phi_{3\lambda}$. If so, our notation would not be too useful. since $\mathbf{C}_1 \oplus \mathbf{C}_2$ would not have a unique result, being $\mathbf{C}_3$ in one case and $\mathbf{C}'_3$ in the second case, where $\mathbf{C}_3$ and $\mathbf{C}'_3$ are different colours.

It is an empirical fact that this difficulty does not arise for additive mixture. $\mathbf{C}_3 \sim \mathbf{C}'_3$ always, under the conditions described. The result leads to the second of Grassmann's Laws.

Note: The preservation of metamerism is a property of additive mixture only. It does not hold for other types of colour mixture, especially subtractive mixture.

3.3.1 Trichromacy

To get an idea of the parameters of metamerism we have to do an experiment. The result of the experiment is the discovery that the visual system is trichromatic ('three-coloured'). The experiment goes as follows:

1. Choose a light of any colour whatsoever.
2. Take a light of a different colour, and, by adjusting its level (brightness) try to match the first colour. This will not be possible unless there has been a fortuitous coincidence in the choice of lights.
3. Take a second light different in colour and mix it additively the first, perhaps by shining it at the same place as the first. Now try to match the initial light by adjusting the level of either or both lights. Once again, the ability to make a match would be fortuitous, unless you happen to be colour-blind (sections 2.1.4 or 3.7.1).
4. Finally, add a third light, and adjust all three of the lights. In most cases you will be able to make a match. (The few where you cannot require subtraction, section 3.3.2.) And, however you change the initial light it is still possible to make a match be readjustment of the three levels.

The fact that exactly three lights are needed for such a match is what is meant when we say that the visual system is trichromatic. That three controls make a match is taken to indicate that the visual system has exactly three degrees of freedom as far as colour is concerned. This fact is usually taken as a consequence of the existence of exactly three cone types.

3.3.2 Subtraction

The experiment described above gives us a natural definition for the subtraction of two stimuli in the context of additive mixture. For each set of three matching lights there are a few test lights which cannot be matched. A match is always possible, however, if addition of matching lights to the test light is allowed. That is, normally a match is specified by

$$\mathbf{C}_t \sim k_1 \cdot \mathbf{C}_1 \oplus k_2 \cdot \mathbf{C}_2 \oplus k_3 \cdot \mathbf{C}_3, \tag{3.10}$$

where the scalar variables k_i indicate the level at which the matching lights are taken for the match. Mixing a light with the test light would produce the match

$$\mathbf{C}_t \oplus k_2 \cdot \mathbf{C}_2 \sim k_1 \cdot \mathbf{C}_1 \oplus k_3 \cdot \mathbf{C}_3, \tag{3.11}$$

if the matching light was number 2. Equation (3.11) could also be written using subtraction $\ominus$ as

$$\mathbf{C}_t \sim k_1 \cdot \mathbf{C}_1 \ominus k_2 \cdot \mathbf{C}_2 \oplus k_3 \cdot \mathbf{C}_3, \tag{3.12}$$

or using a negative level variable k_2 as

$$\mathbf{C}_t \sim k_1 \cdot \mathbf{C}_1 \oplus (-k_2) \cdot \mathbf{C}_2 \oplus k_3 \cdot \mathbf{C}_3. \tag{3.13}$$

Neither subtraction, $\ominus$, nor negative levels can be produced using real lights, since there is no anti-photon. The operations discussed here are purely formal. Match equations written using them always stand for match equations with the subtracted or negative quantities on the other side of the equation.

It is important to note that subtraction, as defined here, has nothing at all to do with subtractive colour mixture (section 3.2.2).

3.3.3 Grassmann's laws

The facts we are currently discussing about the mixing of colours were discovered in the mid-nineteenth century and found concise expression in the form of Grassmann's three laws.

Grassmann's first law

The first law is the expression of trichromacy. It states that if either subtraction or negative levels are allowed that any four colours are linearly dependent. That is, given a set of four colours: $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, and $\mathbf{C}_4$ there exists a set of constants such that

$$\mathbf{C}_4 \sim k_1 \cdot \mathbf{C}_1 \oplus k_2 \cdot \mathbf{C}_2 \oplus k_3 \cdot \mathbf{C}_3. \tag{3.14}$$

This states that the vector space of colours (section 3.3.4) has three dimensions.

Grassmann's second law

Grassmann's second law describes the conservation of metamerism under additive mixture. It is in two parts, the first applying to additive mixture, the second applying to changes of level. The first part is:

$$\begin{array}{ll} \text{If} & \mathbf{C}_1 \sim \mathbf{C}_2 \quad \text{and } \mathbf{C}_3 \sim \mathbf{C}_4, \\ \text{then} & \mathbf{C}_1 \oplus \mathbf{C}_3 \sim \mathbf{C}_2 \oplus \mathbf{C}_4. \end{array} \tag{3.15}$$

Note that this law is trivial when applied to spectral power distributions, showing only that numbers of photons add in a conventional way. It is significant when the $\mathbf{C}_i$'s are colours, since it implies the preservation of metamerism under additive mixture. The second part of the law requires a definition of scalar multiplication. A neutral-density filter attenuates light of all wavelengths equally. If a spectral distribution, Φ_λ, is passed through a neutral density filter with attenuation α then it becomes $\alpha \cdot \Phi_\lambda$, where $0 \leq \alpha \leq 1$. Generalise attenuation to include amplification, and $0 \leq \alpha$. (Note that α cannot be negative because spectral distributions with values less than 0 do not exist.) We call this process scalar multiplication, and it can be extended to the colours $\mathbf{C}$ and $\alpha \cdot \mathbf{C}$ produced by $\Phi(\lambda)$ and $\alpha \cdot \Phi(\lambda)$. The second part is:

$$\begin{array}{ll} \text{If} & \mathbf{C}_1 \sim \mathbf{C}_2, \\ \text{then} & \alpha \cdot \mathbf{C}_1 \sim \alpha \cdot \mathbf{C}_2, \qquad \text{for all} \quad \alpha \geq 0. \end{array} \tag{3.16}$$

Except for the necessity that α be positive these laws produce the definition of a vector space. Section 3.3.2, above, provided the necessary definition of subtraction; section 3.3.4, below, discusses the vector space that can be created from colours.

Grassmann's third law

This law states that the set of all colours is related to the set of all spectral power distributions by a transformation which has no holes, cuts or cliffs. It says that any continuous change in the spectral power distribution of a stimulus Φ_λ produces a continuous change in the perceived colour $\mathbf{C}$.

3.3.4 Vector Space of Colours

Using the definition of subtraction we can assign a meaning to the linear sum $\alpha \cdot P_1 \oplus \beta \cdot P_2$, where α and β are real numbers and P_1 and P_2 are colours. The problem is that not all linear sums represent realisable colours, because some of them can be produced only by spectral distributions with negative values. Thus, it is important to distinguish carefully between two separate concepts. One is the set of all objects $\alpha \cdot P_1 \oplus \beta \cdot P_2$ where α and β are real numbers, P_1 and P_2 are colours. All colours are among these objects; so are an infinite number of non-colours. This set is a vector space; because of trichromacy (section 3.3.1) its dimensionality is three. Let us call this set the vector space of colours. The second is the set of all realisable colours. It is a subset of the vector space of colours, but is not itself a vector space. We will call it the colour solid. (This departs somewhat from normal CIE usage, which reserves the term colour solid for the set of all surface colours, that is, the set of all colours realisable using coloured surfaces. In these notes we will extend the usage to mean all colours realisable using any technology, then modify the usage to indicate the colour generating medium we have in mind, e.g. surface colour solid, monitor colour solid. The colour solid terminology coincides with the established term 'colour gamut'. Thus, the colour gamut of a monitor is the set of all colours that the monitor can produce. Of course, it coincides exactly with the monitor colour solid.) We will develop the properties of the vector space of colours in section 3.4.

3.4 Colour Identification

Metamerism means that the spectral distribution is not a good label for a colour, since many colours that are the same do not have the same label. This section is concerned with presenting the theoretical basis of the labelling scheme used in basic colorimetry. This scheme, which has been in everyday use for about 50 years, is successful in giving each colour a unique label, such that two spectral distributions are seen to be the same colour if and only if they have the same label.

The scheme works by finding a basis for colour space. Then coordinates with respect to the basis can be used as a label for the colour. The definition of the colour space axioms through Grassmann's Laws then ensures that the label is unique. One empirical observation is needed before presenting the theory. It is that the visual system is trichromatic, which has been discussed in sections 3.3.1 and 3.3.3. Trichromacy means that, when subtraction (not subtractive mixing) is allowed, any colour can be matched by a linear combination of at most three arbitrary colours (provided only that the colours are not linearly dependent). This observation amounts to the statement that colour space is three dimensional. We will in fact show that it is isomorphic to $\mathbf{R}^3$. The observation cannot be explained; except to note that it follows logically from the idea that there are exactly three cones, and that the colour seen depends only on the outputs from those cones. It is also worth noting that the success of such devices as colour television monitors in representing a full range of colours using only three components is also evidence for the trichromatic nature of colour vision.

3.4.1 The colour matching functions

Our first task is to develop a basis for colour space, and a method for getting from a spectral distribution to a set of coordinates with respect to the basis. Our development follows the reasoning that was used in the initial nineteenth century research that established these techniques and parallels the experiments that were done to establish trichromacy and metamerism. A stimulus having spectral power distribution Φ_λ can be thought of as the additive mixture of a set of monochromatic (single-wavelength) lights. That is,

$$\Phi_\lambda = \int \Phi_{\lambda'} \delta(\lambda - \lambda')\, d\lambda', \tag{3.17}$$

where $\delta(\lambda - \lambda')$ is the spectral power distribution of a monochromatic stimulus of wavelength λ'. The colour of this stimulus, which we call $\mathbf{C}$, is the additive mixture of a set of colours produced by monochromatic lights. That is,

$$\mathbf{C} \sim \oint \Phi'_\lambda \mathbf{1}_{\lambda'}\, d\lambda', \tag{3.18}$$

where '$\oint$' is the integral equivalent of additive mixture, and $\mathbf{1}_\lambda$ is the colour produced by a 'unit' stimulus of wavelength λ.

Remember that earlier we said that any light can be matched by the additive mixture of three lights. We choose three arbitrary standard lights to use for matching. They are called primaries, and are indicated by $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$. Then we can match each of the monochromatic colours by an additive mixture of the three primaries:

$$\mathbf{1}_\lambda \sim \bar{\pi}_1(\lambda) \cdot \mathbf{P}_1 \oplus \bar{\pi}_2(\lambda) \cdot \mathbf{P}_2 \oplus \bar{\pi}_3(\lambda) \cdot \mathbf{P}_3. \tag{3.19}$$

Equation (3.19) states that any monochromatic light can be matched by appropriate amounts of the three standard lights. $\bar{\pi}_1(\lambda)$, $\bar{\pi}_2(\lambda)$, and $\bar{\pi}_3(\lambda)$, are determined experimentally, by having an observer vary the amounts of the three primaries until he sees the combination match the monochromatic stimulus. They depend on wavelength, and may be negative, which means that for the wavelength where the function is negative a match is achieved by adding the standard to the monochromatic stimulus, instead of to the combination of matching lights. Since the spectral power distribution, Φ_λ is the sum of monochromatic distributions, the colour matches the additive mixture of a sum of colours produced by monochromatic lights.

$$\begin{aligned} \mathbf{C} &\sim \int \Phi_\lambda(\bar{\pi}_1(\lambda)\mathbf{P}_1 \oplus \bar{\pi}_2(\lambda)\mathbf{P}_2 \oplus \bar{\pi}_3(\lambda)\mathbf{P}_3)\, d\lambda \\ &\sim P_1 \cdot \mathbf{P}_1 \oplus P_2 \cdot \mathbf{P}_2 \oplus P_3 \cdot \mathbf{P}_3 \end{aligned} \tag{3.20}$$

where

$$P_i = \int \Phi_\lambda \bar{\pi}_i(\lambda)\, d\lambda; \qquad \text{for} \quad i = 1, 2, 3. \tag{3.21}$$

In vector space language, the set $\{\mathbf{P}_1,\ \mathbf{P}_2,\ \mathbf{P}_3\}$ is a basis for colour space and the triplet of numbers (P_1, P_2, P_3) is the set of coordinates of colour $\mathbf{C}$ with respect to that basis. Colour science has its own names for these concepts. The set of lights $\{\mathbf{P}_1,\ \mathbf{P}_2,\ \mathbf{P}_3\}$ is called a set of primaries, or primary stimuli, and the numbers (P_1, P_2, P_3) are the tristimulus values of the stimulus $\mathbf{C}$ with respect to the primaries. The set of functions from which tristimulus values are derived is known as a set of colour matching functions.

3.4.2 Tristimulus values

The following properties of tristimulus values, colour matching functions, and primaries are important:

1. Primaries are chosen conventionally, and need not be monochromatic. In fact, they need not even be realizable as spectral distributions.
2. Colour matching functions are derived empirically from primaries via colour matching experiments.
3. The tristimulus values of a colour have a conventional element (from the primaries) and an empirical element (from the colour matching functions).
4. When a linear combination of two colours is produced by additive mixture

   $$\mathbf{C}_3 \sim \alpha \cdot \mathbf{C}_1 \oplus \beta \cdot \mathbf{C}_2, \tag{3.22}$$

   the tristimulus values of the resulting colour are the same linear combination of the tristimulus values,

   $$P_{3i} = \alpha P_{1i} + \beta P_{2i}. \tag{3.23}$$

   Note that while the colour addition uses the additive mixture notation, the tristimulus values, being numbers, use conventional addition and multiplication.
5. Because additive mixture with positive α and β is always possible, the colour solid must be a convex subset of colour space. Thus, in the $\mathbf{R}^3$ space with tristimulus value coordinates, if (P_{11}, P_{12}, P_{13}) are the tristimulus values of a physical colour (part of the colour solid), and (P_{21}, P_{22}, P_{23}) are the tristimulus values of a second physical colour, then all sets of tristimulus values on the straight line joining them correspond to physical colours.
6. Primaries in common use are generally chosen so that all tristimulus values of all colours in the colour solid are positive. In other words, the colour solid lies in the all-positive octant with respect to these bases.

3.5 Change of Primaries

The set of primaries which are used may be chosen at the user's convenience. Changing from one colour measuring system to another or from one colour generating device to another (assuming that the colour generating devices are additive as a colour monitor is) often involves not much more than changing from one set of primaries to another. Fortunately, changing from one set of primaries to another is very simple in the mathematical formalism we have chosen. Let the new set of primaries be $\{\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3\}$. The colour matching functions derived from them are $\bar{\varrho}_1(\lambda)$, $\bar{\varrho}_2(\lambda)$, and $\bar{\varrho}_3(\lambda)$. The tristimulus values of colour C in this new system are

$$Q_i = \int \Phi_\lambda \bar{\varrho}_i(\lambda)\, d\lambda \tag{3.24}$$

What are the colour matching functions of the new set of primaries? How are they related to the colour matching functions of the old set of primaries? And what is the relationship between the two sets of tristimulus values? To answer questions like these we must know the relationship between $\bar{\pi}_i(\lambda)$ and $\bar{\varrho}_i(\lambda)$. To get it note that each of the new primaries can be treated as a stimulus with a spectral power distribution (which may or may not be physical). Thus there are three matching equations:

$$\begin{aligned}
\mathbf{Q}_1 &\sim A_{11} \cdot \mathbf{P}_1 \oplus A_{12} \cdot \mathbf{P}_2 \oplus A_{13} \cdot \mathbf{P}_3 \\
\mathbf{Q}_2 &\sim A_{21} \cdot \mathbf{P}_1 \oplus A_{22} \cdot \mathbf{P}_2 \oplus A_{23} \cdot \mathbf{P}_3 \\
\mathbf{Q}_3 &\sim A_{31} \cdot \mathbf{P}_1 \oplus A_{32} \cdot \mathbf{P}_2 \oplus A_{33} \cdot \mathbf{P}_3
\end{aligned} \tag{3.25}$$

which we can express in matrix notation as

$$\mathbf{Q}_i \sim \bigoplus_{j=1}^{3} A_{ij} \cdot \mathbf{P}_j, \tag{3.26}$$

with

$$A_{ij} = \int \Psi_{i\lambda} \bar{\pi}_j(\lambda)\, d\lambda, \tag{3.27}$$

where $\Psi_{i\lambda}$ is the spectral distribution of $\mathbf{Q}_i$. Note that the the entries A_{ij} in the matrix are the tristimulus values of the new primaries with respect to the old ones. Now, take the colour matching equation,

$$\mathbf{1}_\lambda \sim \bigoplus_{i=1}^{3} \bar{\varrho}_i(\lambda) \cdot \mathbf{Q}_i, \tag{3.28}$$

which defines the new colour matching function, substitute equation (3.25) for the new set of primaries, and compare it to the definition of the old colour matching functions, equation (3.19).

$$\begin{aligned}
\mathbf{1}_\lambda &\sim \bigoplus_{j=1}^{3} \sum_{i=1}^{3} \bar{\varrho}_i(\lambda) A_{ij} \cdot \mathbf{P}_j \\
&\sim \bigoplus_{j=1}^{3} \bar{\pi}_j(\lambda) \cdot \mathbf{P}_j.
\end{aligned} \tag{3.29}$$

Since there are three primaries the sums are unique, and we can make identify the relationship of the old and new colour matching functions in equation (3.29). It is

$$\bar{\pi}_j(\lambda) = \sum_{i=1}^{3} \bar{\varrho}_i(\lambda) A_{ij}. \tag{3.30}$$

The inverse matrix A^{-1} makes the inverse transformation from the new to the old colour matching functions,

$$\bar{\varrho}_i(\lambda) = \sum_{j=1}^{3} \bar{\pi}_j(\lambda) A^{-1}_{ij}. \tag{3.31}$$

We get the tristimulus values in the new system using

$$Q_i = \sum_{j=1}^{3} A_{ij}^{-1} P_j. \tag{3.32}$$

To sum up, the matrix A, of which the entries are the tristimulus values of the new primaries in terms of the old colour matching functions, can be used to go from the new system to the old one:

$$\bar{\pi}_j(\lambda) = \sum_{i=1}^{3} \bar{\varrho}_i(\lambda) A_{ij}, \tag{3.30}$$

and

$$P_i = \sum_{j=1}^{3} A_{ij} Q_j. \tag{3.33}$$

Its inverse is used to go from the old system to the new one:

$$\bar{\varrho}_i(\lambda) = \sum_{j=1}^{3} \bar{\pi}_j(\lambda) A_{ij}^{-1}. \tag{3.31}$$

and

$$Q_i = \sum_{j=1}^{3} A_{ij}^{-1} P_j. \tag{3.32}$$

This excursion into the technicalities of primaries and colour matching functions may seem overly specialised, but it forms the basic theory of the CIE system and of calibration technology. The CIE colorimetric system, as well as several calibration techniques, are special cases of this general formalism.

3.6 Chromaticity and Other Colour Terms

Now we can present some important colour concepts in terms of our general colour labelling scheme. These concepts will be treated in the context of a particular colour system when the CIE colorimetric system is discussed below.

3.6.1 Chromaticity coordinates

Tristimulus values have the property that each coordinate is multiplied by the same constant when a light changes in level without its relative spectral distribution changing. A good example of this process occurs with video monitor phosphors which, if the monitor is well-aligned, change their light output without changing the relative spectral distribution when the gun voltage changes. When level changes of this kind occur, colour appearance does not change very much; our perception of hue and saturation remains more or less true, for example, when lighting levels change. Thus, it is useful to have a colour representation which isolates degrees of freedom orthogonal to overall level change and presents them in a plane. (This puts what seem to be the two most salient aspects of colour, hue and saturation, in a plane, which is easy to illustrate and comprehend.) The conventional, and now widely accepted, method for doing so forms chromaticity coordinates, which are defined in terms of tristimulus values:

$$\begin{aligned} p_1 &= P_1/(P_1 + P_2 + P_3) \\ p_2 &= P_2/(P_1 + P_2 + P_3) \\ p_3 &= P_3/(P_1 + P_2 + P_3) \end{aligned} \tag{3.35}$$

Four properties of the chromaticity coordinates are immediately obvious:

1. They are independent of level. Multiplying P_1, P_2 and P_3 by the same constant does not change the p's.

2. They sum to 1; only two are independent. Thus we can plot them on a plane.
3. They are not defined when no light is present (i.e. $P_1 = P_2 = P_3 = 0$).
4. It is usual to choose a set of primaries that makes all tristimulus values non-negative for all physically possible spectral distributions. When that is true all chromaticity coordinates lie between 0 and 1.

$$0 \leq p_i \leq 1. \tag{3.36}$$

This means that all colours are represented in a finite area.

Three properties of chromaticity coordinates that are less obvious concern additive mixture.

5. When two colours are mixed additively the chromaticity coordinates of the resulting colour lie on the straight line joining the chromaticity coordinates of the two original colours. The actual coordinates cannot, however, be produced by a linear interpolation scheme, as the tristimulus values can be.
6. The set of physically realisable colours - the whole colour solid projected into the chromaticity plane - is a convex region.
7. The colour solid, or colour gamut, of any device has an intersection with the chromaticity plane (at any given luminance). This intersection may or may not be the same at different luminances. For a video monitor, or for any other device which creates its colours by additive mixture, this intersection has two simple properties:
    1. It is independent of luminance up to the lowest luminance where the guns (primaries) saturate (reach their maximum output).
    2. It is a polygon, whose corners are the chromaticites of the guns (primaries of the colour generating device).

    Thus, the colour solid (gamut) of a monitor is a triangle in chromaticity coordinates, and the corners of the triangle are the chromaticity coordinates of the phosphors.

3.6.2 The spectrum locus

We would like to know what the boundary of the colour solid is. The answer, as it turns out, is quite simple: the boundary of the colour solid consists of the tristimulus values of monochromatic lights, plus those that lie on the purple boundary (section 3.6.3). Let us look first at tristimulus values of monochromatic lights which, taken together, are called the spectrum locus. This surface is somewhat difficult to visualize. To do so first note that the tristimulus values of a monochromatic light of variable intensity lie on a ray extending from the origin. Then, note the empirical fact that the tristimulus values of monochromatic lights between 560nm (yellow-green) and 700nm (red) lie on a plane. Finally note that the tristimulus values of monochromatic lights between 560nm and 400nm (blue) lie on a convex curving surface. We know the surface to be convex because no monochromatic light in this region can be matched by combinations of other monochromatic lights unless some of the matching lights are subtracted (i.e. added to the light between 560nm and 400nm.) All points that lie within this surface are parts of the colour solid because they can be made as an additive combination of two monochromatic lights. Points lying outside the surface (i.e. on projections of lines crossing through the enclosed volume) do not correspond to physically realizable colours because their production requires a negative amount of the monochromatic light directly across the surface from them. Figure 3.2 is an attempt schematically to illustrate the spectrum locus in the space of tristimulus values.

3.6.3 The purple boundary

Light of wavelength longer than 700 nm is not visible, nor is light of wavelength shorter than 400 nm. Yet, the tristimulus values of these two lights do not coincide (since they are not the same colour; 700 nm is deep red, while 400 nm is violet or deep blue). Thus, the surface produced by the spectrum locus is not closed, and the surface that completes the limits of the colour solid is made up of additive mixtures of 400 nm and 700 nm. These colours, which are mixtures of red and blue, are called the purple boundary. Both the spectrum locus and the purple boundary are easier to visualise in terms of chromaticity coordinates.

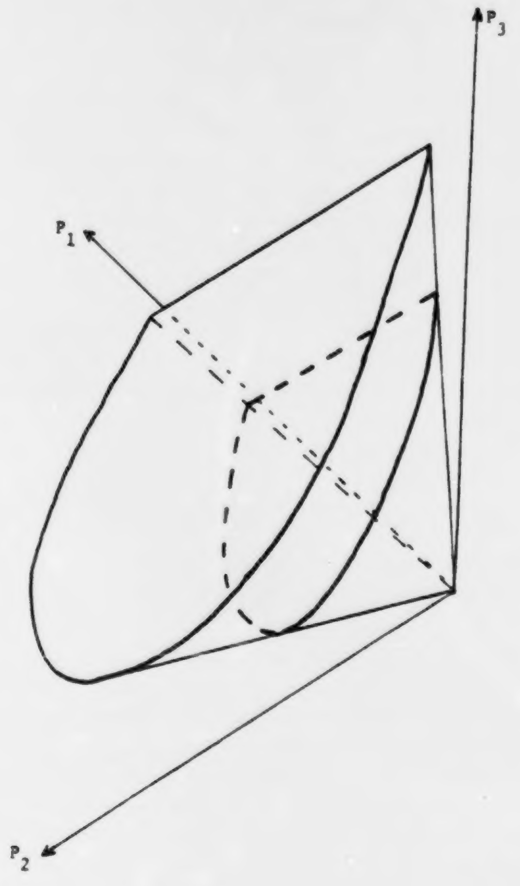

Figure 3.2. The colour solid, shown in a space of arbitrary tristimulus values. It is a cone extending from the origin, having a roughly horseshoe-shaped cross-section. The connection between this shape and the spectrum locus in chromaticity coordinates is not easy to see. The best way is to turn the diagram to make P_1 horizontal, and then to try to envision the projection of the cross-section onto the P_1–P_2 plane.

3.6.4 The spectrum locus in chromaticity coordinates

See figure 3.3. The chromaticity coordinates of spectral colours from 560 nm (yellowish-green) to 700 nm (deep red) lie in a straight line, often called the Rayleigh line. Those of colours from 400 nm (deep blue) to 560 nm lie on a convex curve. The whole curve, which is the spectrum locus in chromaticity coordinates, lies within the box defined by $0 \leq p_1 \leq 1$, $0 \leq p_2 \leq 1$. Joining the point at 400 nm to the point at 700 nm is the purple boundary. Points within the figure are the chromaticity coordinates of colours that can be realised using real lights (i.e. of the colour solid). Points outside the figure are not real colours, and could be produced only by spectral distributions that are negative at some wavelengths . This discussion has treated the properties of general tristimulus values and chromaticity coordinates, defined in terms of arbitrary primary lights, in some detail. We will return to this topic later when the CIE system of colorimetry is discussed. At that point the properties of a particular set of primaries will be explained. A schematic illustration of the spectrum locus in arbitrary chromaticity coordinates (also called the chromaticity diagram) is shown in figure 3.3.

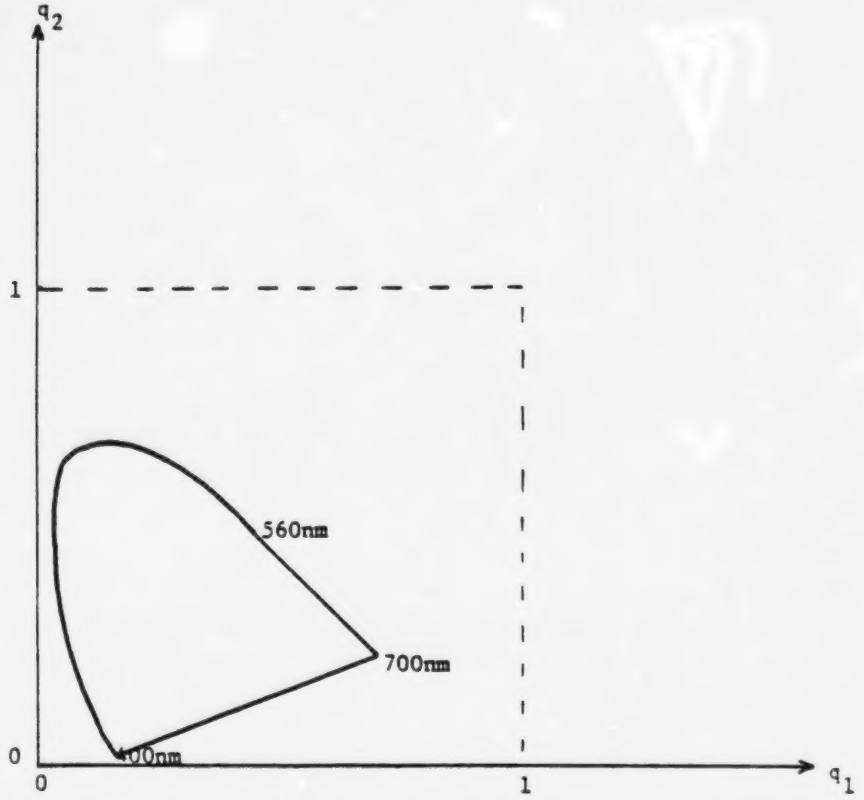

Figure 3.3. Projection of the colour solid onto arbitrary chromaticity coordinates. The curved line is the spectrum locus, with several wavelengths marked on it; the straight line joining 400 nm to 700 nm is the purple boundary.

3.7 Limits of Basic Colorimetry

We now end our discussion by describing a few of the limitations of the colorimetric formalism that we have developed here. They should be regarded as unsolved problems, currently the subject of research in the colorimetric community. At present we do not know solutions will be provided by simple extensions of the present formalisms or whether entirely new methods will be needed to solve them successfully.

3.7.1 Colour difference

The methods of this section make it possible to write a specification for two colours to be exactly the same. In the real world, of course, nothing is ever exact, and a specification would have to include tolerances indicating how much error can be neglected. Presumably, such an error specification should be based on human perception: we would want to tolerate an amount of mismatch that is negligible to a human observer. But, there is no simple relationship between tristimulus values of two colour stimuli and the colour difference perceived between them. Two specifications recommended by the CIE for such a purpose do exist: CIELAB and CIELUV, which are described in lecture 15. These specifications are based on colour difference data taken under a restricted range of viewing conditions, and are not based on a fundamental understanding of the perception of colour difference, as is the basic colorimetry described in this lecture. Thus, it is not clear how well they apply to other types of task or viewing condition. This, then, along with several other visual functions, such as brightness, is an area where there are standards known to be useful in some applications, but where the use of the standards is not underwritten by basic understanding of the relevant visual processes. Basic colorimetry does not answer such questions.

3.7.2 Colour appearance

Another area where basic colorimetry has no answers is colour appearance. Basic colorimetry may assure us that two colours are the same colour, but it does not tell us what colour they appear to a human observer. Consider two orange stimuli viewed in dark surrounds. Colorimetry tells us that they are the same and so they appear. Change the surround to a white one. The stimuli now appear to be brown. Their colorimetric coordinates have not changed, so they still appear to be the same as one another, but their appearance has changed qualitatively. This phenomenon, qualitative changes in appearance without change in colorimetric coordinates is well known and shows that colorimetry cannot be used as a measure of colour appearance. This is another area of active research.

3.7.3 Colour defective observers

Most of the above remarks are correct in general but wrong in detail when applied to the colour defective observers described in section 2.1.4. Those observers are thought to lack a single dimension of colour response. On that view, all the discussion above continues to hold, except that colour space is two-dimensional, isomorphic to $\mathbf{R}^2$. For such an observer the colour solid is an area; tristimulus values are two numbers; and the chromaticity diagram is a line. The colour defective observer would be identified by his ability to make a good match to any colour using an additive mixture of only two primaries, and schemes have been developed to identify and classify different types of colour defect. Extensive discussion of them would go well beyond the intended scope of the present notes.

3.8 References

Wyszecki, G. and Stiles, W. S., 1982, *Color Science*, second edition, New York: Wiley.

Notes

Notes

Lecture 4: A. R. Robertson

THE CIE SYSTEM

4.1 Introduction

The recognized international authority on photometry and colorimetry is the International Commission on Illumination. This is known as the CIE from the initials of its French name, Commission Internationale de l'Eclairage. The various definitions and procedures recommended by the CIE are the subject of this lecture.

All the recommendations are given in CIE Publication No. 18.2 (1983) for photometry and No. 15.2 (1986) for colorimetry. A thorough treatment of the concepts involved and their application is given in several standard textbooks such as Billmeyer and Saltzman (1981), Judd and Wyszecki (1975), Wright (1969), and Wyszecki and Stiles (1982).

4.2 Photometry

4.2.1 Radiometric and photometric units

There are several quantities that are used to describe radiant energy quantitatively. Each one applies to a different geometrical situation. For each radiant quantity there is an equivalent luminous quantity to describe the eye's response. The most important radiant quantities are listed in Table 4.1 together with the luminous equivalents and the most commonly used symbols. A subscript e on any symbol indicates a radiant quantity, and a subscript v indicates a luminous quantity. These quantities will be discussed in more detail in lecture 7.

Radiometric Quantity	Unit	Luminous Quantity	Unit	Symbol
Radiant flux or power	watt	Luminous flux	lumen	Φ or P
Radiant intensity	watt per steradian	Luminous intensity	lumen per steradian or candela	I
Irradiance	watt per square metre	Illuminance	lumen per square metre or lux	E
Radiant exitance	watt per square metre	Luminous exitance	lumen per square metre	M
Radiance	watt per steradian per square metre	Luminance	candela per square metre	L

Table 4.1. Names, units and symbols of the most important radiometric and photometric quantities.

4.2.2 $V(\lambda)$ and $V'(\lambda)$ curves

The CIE has standardized two weighting functions by which luminous quantities (such as luminance) can be calculated from radiant quantities (such as radiance). These are known as the spectral luminous efficiency functions for photopic and scotopic vision and are abbreviated $V(\lambda)$ and $V'(\lambda)$ respectively. They are shown in figure 4.1. The luminous quantities are calculated by equations such as:

$$L_v = K_m \int L_{e\lambda} V(\lambda)\, d\lambda,$$

and

$$L'_v = K'_m \int L_{e\lambda} V'(\lambda)\, d\lambda.$$

Photopic vision is mediated essentially or exclusively by the cones in the human retina and functions at normal daylight levels. Scotopic vision is mediated essentially or exclusively by the rods and functions at very low levels. There is no colour vision at scotopic levels. Vision at intermediate levels, where both cones and rods are active,is known as mesopic vision.

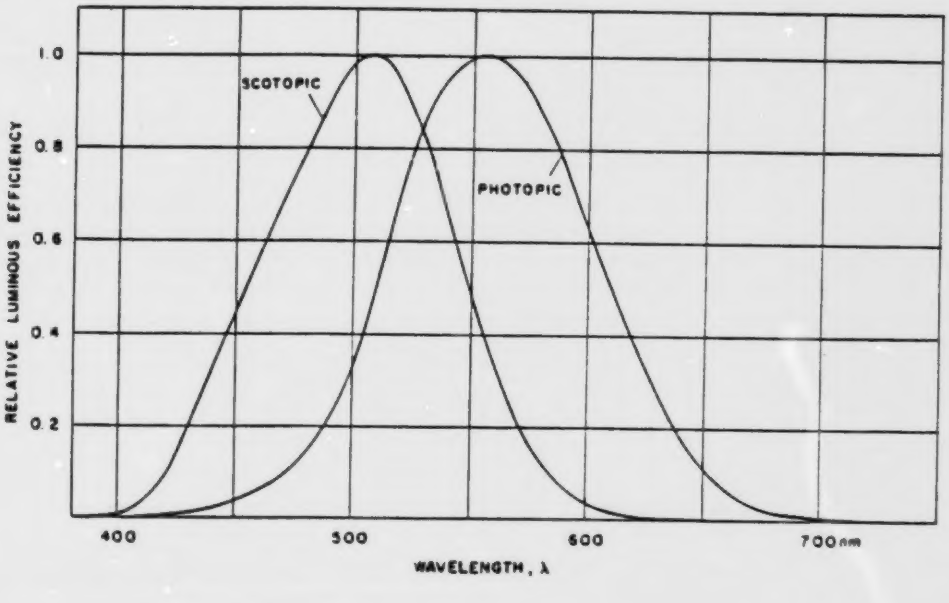

Figure 4.1. CIE standard relative luminous efficiency functions for photopic vision ($V(\lambda)$) and scotopic vision ($V'(\lambda)$). (Wyszecki and Stiles, 1982).

4.2.3 Definition of the candela

The basic unit of photometry, the candela,is defined as the luminous intensity of a monochromatic source having a radiant intensity of 1/683 watts per steradian at 555 nm. By setting the constants K_m and K'_m in the above equations to:

$K_m = 683$ photopic lumens/watt
= maximum spectral luminous efficacy of monochromatic radiation for photopic vision,

and

$K'_m = 1700$ scotopic lumens/watt
= maximum spectral luminous efficacy of monochromatic radiation for scotopic vision,

we have

$$K_m V(555\mathrm{nm}) = K'_m V'(555\mathrm{nm}) = 683 \quad \mathrm{lm}\cdot\mathrm{w}^{-1},$$

in accordance with the definition.

4.2.4 Typical luminances

Snow on clear mid-winter day	$30,000$ cd · m^{-2}
White sheet in operating theatre	$5,000$ cd · m^{-2}
White paper on office desk	200 cd · m^{-2}
Floor of freight elevator	5 cd · m^{-2}
White paper in moonlight	50×10^{-3} cd · m^{-2}
White paper at night, no moon	0.3×10^{-3} cd · m^{-2}

Very approximately: $L_v > 3$ cd · m^{-2} is photopic
$L_v < 10^{-3}$ cd · m^{-2} is scotopic

The range between these is mesopic.

4.3 Colour Matching

In a simple colour-matching experiment, as shown in figure 4.2, a test light is matched by a mixture of three primary lights. Three are necessary and sufficient because human eyes have three types of cones (except for 'colour-blind' people who have only two types).

As explained in lecture 3, the amounts of the three primaries needed to make the match are called tristimulus values and provide a means of specifying the colour of the test stimulus. The match is called a 'metameric' match because even though the two lights (the test light and the mixture of the primaries) have the same colour and appear to be identical they are in fact different in their physical composition.

Sometimes, in order to make a match, it is necessary to mix one of the primaries with the test light and to match this mixture with a mixture of the other two primaries. The amount of the first primary is then taken to be negative.

4.3.1 Monochromatic lights

Lights containing energy at only a single wavelength are called monochromatic lights.

4.3.2 Colour-matching functions

Tristimulus values of monochromatic lights of equal energy are called colour-matching functions. One set of colour-matching functions, in terms of a particular set of red, green and blue monochromatic primaries, is shown in figure 4.3.

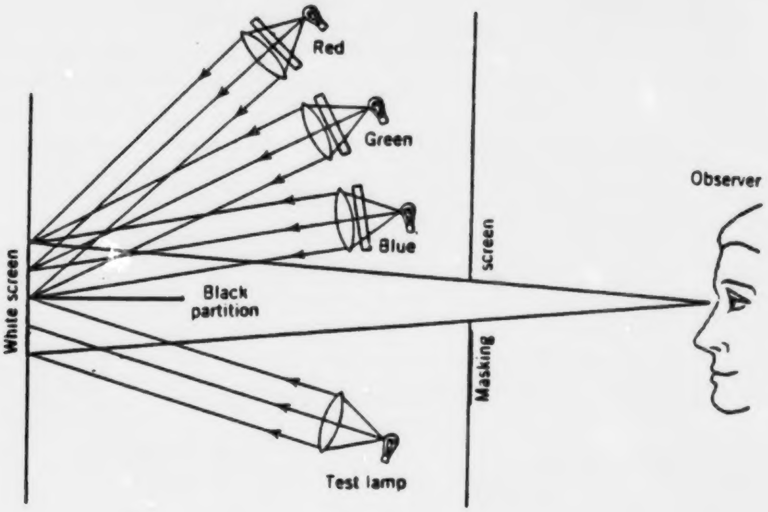

Figure 4.2. Simple colour-matching experiment (Billmeyer and Saltzman, 1981).

The tristimulus values of a light containing energy at more than one wavelength are given by

$$R = \int L_{e\lambda} \bar{r}(\lambda)\, d\lambda,$$

$$G = \int L_{e\lambda} \bar{g}(\lambda)\, d\lambda,$$

and

$$B = \int L_{e\lambda} \bar{b}(\lambda)\, d\lambda.$$

4.3.3 Transformation of primaries

Tristimulus values in terms of one set of primaries can be transformed to any other set by three linear equations:

$$\begin{aligned} X &= a_{11}R + a_{12}G + a_{13}B, \\ Y &= a_{21}R + a_{22}G + a_{23}B, \\ Z &= a_{31}R + a_{32}G + a_{33}B. \end{aligned}$$

The absorptions in the three types of cones are a set of tristimulus values. The primaries are lights that stimulate each of the types without stimulating the other two. Because of the overlapping spectral responsivities of the cone types, no real lights can do this so the primaries are imaginary.

Another set of imaginary primaries is the basis of the CIE system of colorimetry. It was chosen for several practical reasons, the main one being to avoid negative colour-matching functions. The resulting colour-matching functions are shown in figure 4.4 and define the CIE 1931 Standard Observer. They are the average of measurements made on a number of real observers.

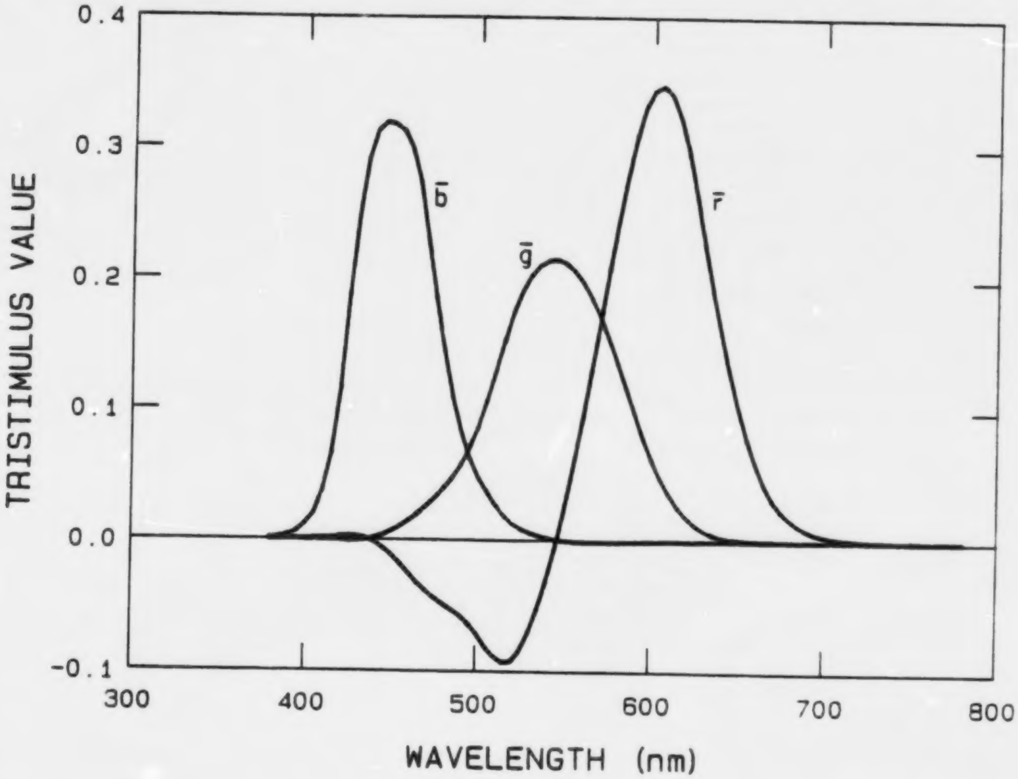

Figure 4.3. Tristimulus values of monochromatic lights of equal energy measured in terms of red, green and blue monochromatic primaries.

4.4 CIE Specifications for Colour-Stimuli

4.4.1 CIE tristimulus values

CIE tristimulus values are calculated by:

$$X = K_m \int L_{e\lambda} \bar{x}(\lambda)\, d\lambda,$$

$$Y = K_m \int L_{e\lambda} \bar{y}(\lambda)\, d\lambda,$$

and

$$Z = K_m \int L_{e\lambda} \bar{z}(\lambda)\, d\lambda.$$

The $\bar{y}(\lambda)$ function is identical to $V(\lambda)$ so the Y tristimulus value is equal to the luminance. If $L_{e\lambda}$ is replaced by another radiometric quantity, Y becomes the corresponding luminous quantity.

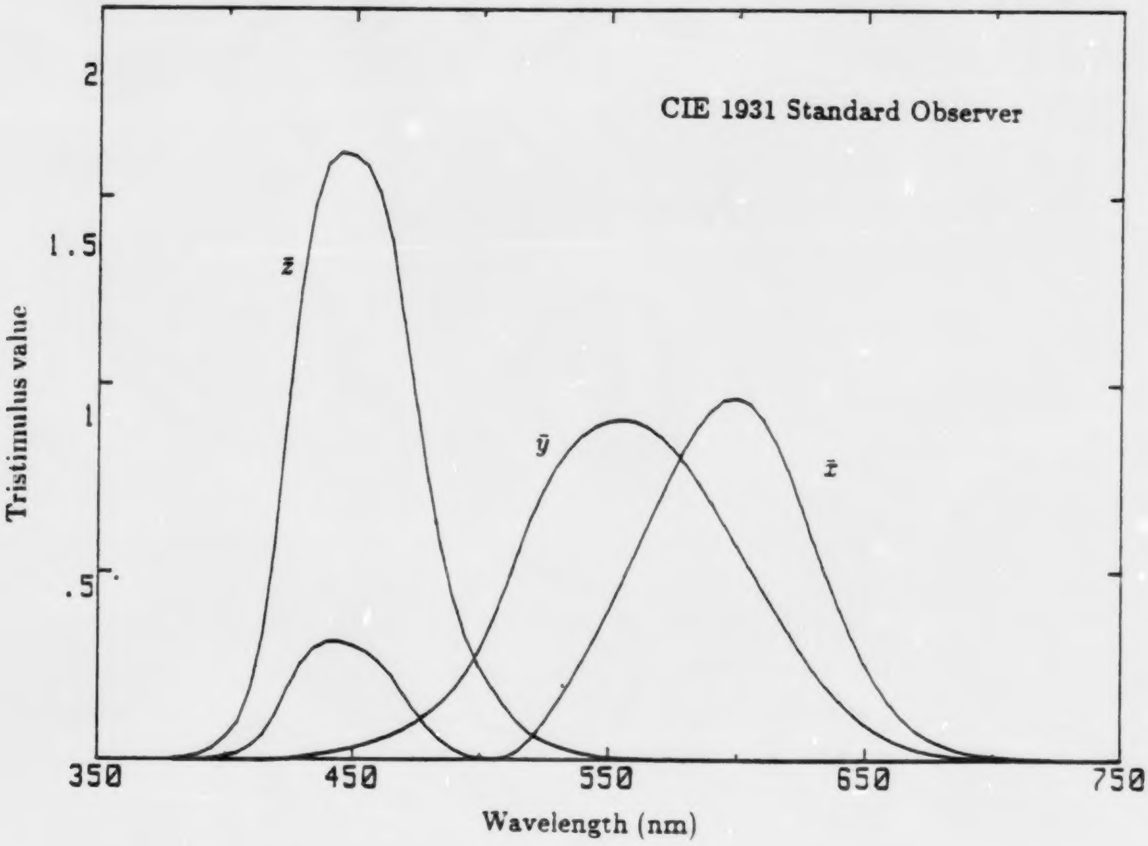

Figure 4.4. Colour-matching functions of the CIE 1931 Standard Observer.

4.4.2 CIE chromaticity coordinates

It is often convenient to calculate two new quantities, called chromaticity coordinates:

$$x = X/(X + Y + Z),$$

and

$$y = Y/(X + Y + Z),$$

where X, Y, and Z are tristimulus values.

The diagram formed by plotting x and y in rectangular coordinates is known as the CIE 1931 chromaticity diagram (figure 4.5). When monochromatic lights are plotted in this diagram they form a horseshoe-shaped curve known as the spectrum locus. The line joining the ends of the spectrum locus is known as the purple boundary. All real colours lie within the area enclosed by the spectrum locus and purple boundary.

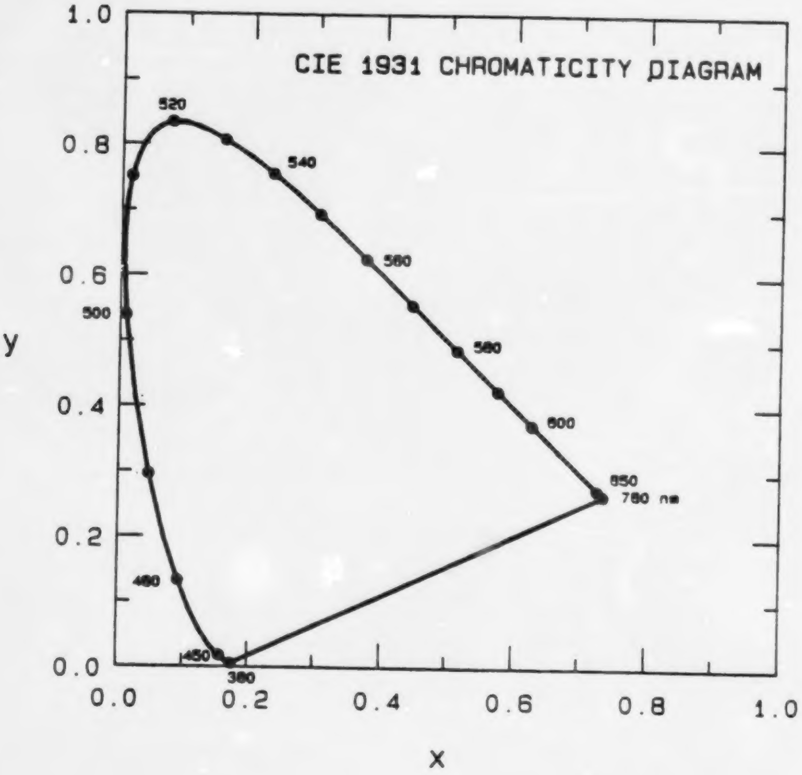

Figure 4.5. CIE 1931 chromaticity diagram showing the spectrum locus and the purple boundary.

4.4.3 Dominant wavelength and excitation purity

The dominant wavelength of a colour stimulus is defined as the wavelength of the monochromatic light that, when combined in suitable proportions with a specified achromatic (white) light, yields a match with the stimulus considered.

For colour stimuli not possessing a dominant wavelength, it is always possible to match the specified achromatic stimulus with a mixture of the given stimulus and a monochromatic stimulus. The wavelength of the latter is called the complementary wavelength of the stimulus considered.

Excitation purity is the distance in the chromaticity diagram from the specified achromatic light to the colour considered, relative to the distance from the achromatic point to the spectrum locus or the purple boundary in the same direction.

The concepts of dominant wavelength, complementary wavelength and excitation purity are illustrated in figure 4.6.

4.4.4 Correlated Colour Temperature

Before the widespread use of fluorescent and other discharge lamps, most light sources were incandescent and therefore had a colour similar to a blackbody radiator. Because of this it became common to describe the colour of a light source by its 'correlated colour temperature'. This is the temperature of the blackbody radiator whose colour is closest to that of the light considered.

As the temperature of a blackbody radiator is increased, its chromaticity coordinates move along a curved line in the chromaticity diagram as shown in figure 4.7. Obviously if the colour of a light source is very far from this line, correlated colour temperature is not a very good descriptor of its colour.

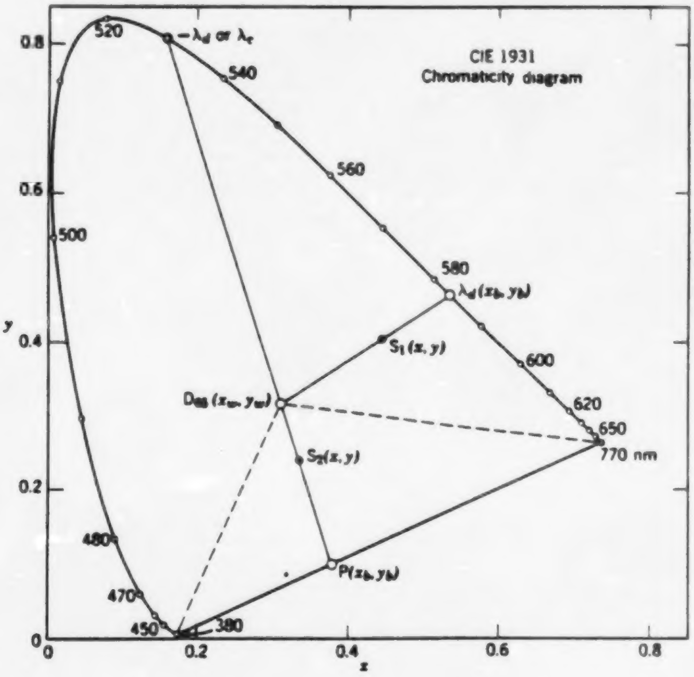

Figure 4.6. CIE 1931 chromaticity diagram showing the concepts of dominant wavelength (λ_d) for stimulus S_1 and complementary wavelength (λ_c) for stimulus S_2 (Wyszecki and Stiles, 1982).

4.4.5 Perceptual terms

The purpose and main use of the CIE system is to predict metameric matches. Two lights with the same tristimulus values will always match if the conditions under which they are viewed is the same. The system is not designed to predict appearance, but there is an approximate correlation as shown in table 4.2. This correlation is dependent on a number of factors, particularly other lights viewed simultaneously with or prior to the one considered.

Perceptual quantities (subjective)	Psychophysical quantities (objective)
Hue	Dominant wavelength
Saturation	Excitation purity
Brightness	Luminance
Lightness	Luminous reflectance or Luminous transmittance

Table 4.2. Approximate correlation between perceptual and psychophysical quantities.

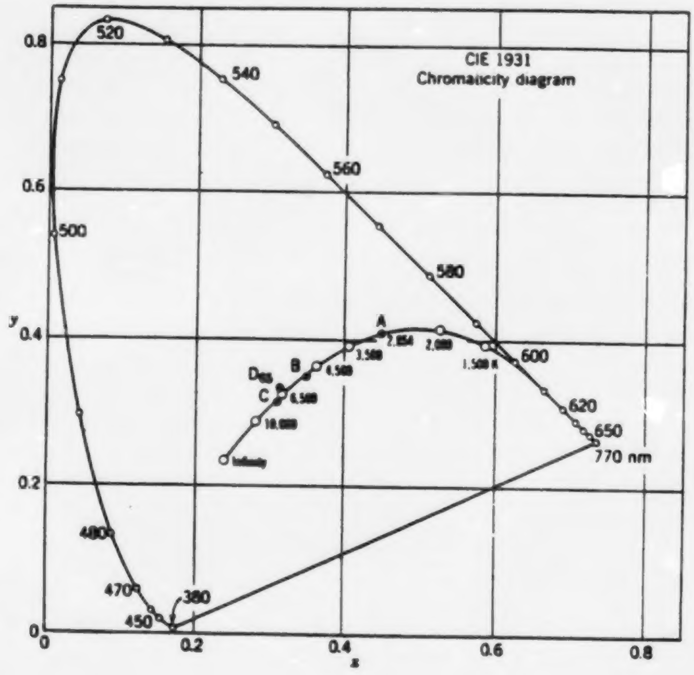

Figure 4.7. CIE 1931 chromaticity diagram showing the locus of blackbody radiators and standard illuminants A, B, C and D_{65} (Judd and Wyszecki, 1975).

4.5 CIE Specifications for Reflecting and Transmitting Materials

For a reflecting specimen it is customary to normalize tristimulus values in the following way:

$$X = k \int R(\lambda)S(\lambda)\bar{x}(\lambda)\,d\lambda,$$

$$Y = k \int R(\lambda)S(\lambda)\bar{y}(\lambda)\,d\lambda,$$

and

$$Z = k \int R(\lambda)S(\lambda)\bar{z}(\lambda)\,d\lambda,$$

where

$$k = 100 / \int S(\lambda)\bar{y}(\lambda)\,d\lambda.$$

$R(\lambda)$ is the spectral reflectance factor and $S(\lambda)$ is the relative spectral power distribution of the source. This normalization ensures that a perfectly white object ($R(\lambda) = 1.0$ at all wavelengths) will always have a Y tristimulus value of 100. This correlates with the fact that perceived colours of a white object remain approximately constant regardless of the level of illumination.

For transmitting specimens, $R(\lambda)$ is replaced by the transmittance or transmittance factor ($T(\lambda)$) but otherwise the same considerations apply.

4.5.1 Standard illuminants

In order to facilitate the comparison of data, the CIE recommends several standard spectral power distributions $S(\lambda)$ for use in calculating the tristimulus values of reflecting and transmitting materials. These distributions are called 'standard illuminants' to distinguish them from 'standard sources' that may or may not be good physical representations of them.

The most important CIE standard illuminants are known as illuminants A and D_{65}. The relative spectral power distributions of standard illuminants A, B, C and D_{65} are shown in figure 4.8. Their chromaticity coordinates are shown in figure 4.7.

CIE standard illuminant A

Illuminant A is a standard illuminant representing light from a blackbody radiator at temperature 2856K. It is defined by a tabulated relative spectral power distribution.

CIE standard illuminant D_{65}

Illuminant D_{65} is a standard illuminant representing a phase of daylight with a correlated colour temperature of 6500K. It is defined by a tabulated relative spectral power distribution.

CIE standard illuminants B *and* C

Standard illuminants B and C are old representations of daylight that are now gradually being superseded by standard illuminant D_{65}.

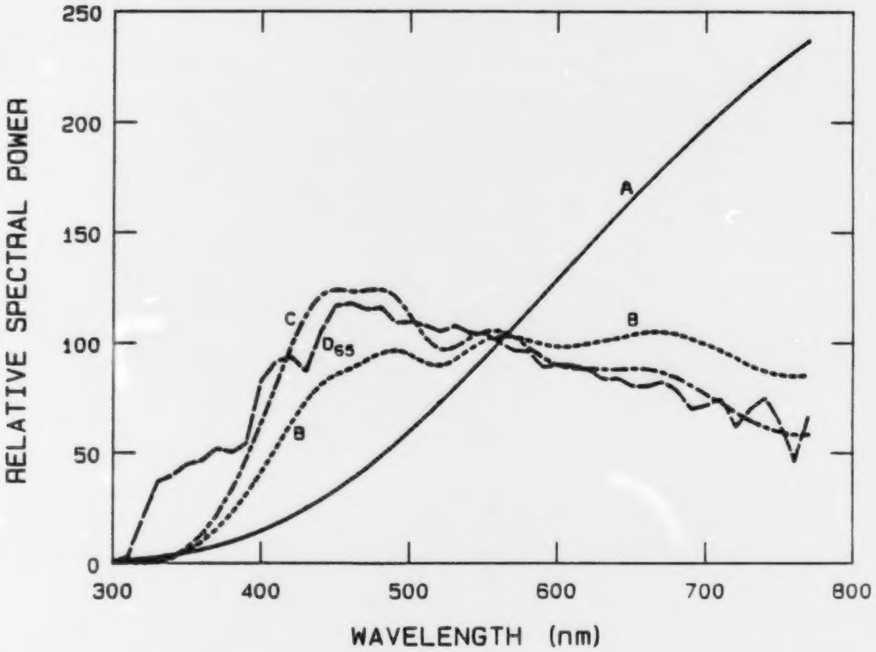

Figure 4.8. Relative spectral power distributions of CIE standard illuminants A, B, C and D_{65}.

4.6 CIE 1964 Standard Observer

In 1964 the CIE added another standard observer to its recommendations. The colour-matching functions of this observer are shown in figure 4.9. They are recommended for use for stimuli subtending an angle of more than 4° at the eye. The 1964 standard observer is based on colour matching in a 10° field so it is often known as the 10° observer. Similarly the 1931 standard observer is known as the 2° observer.

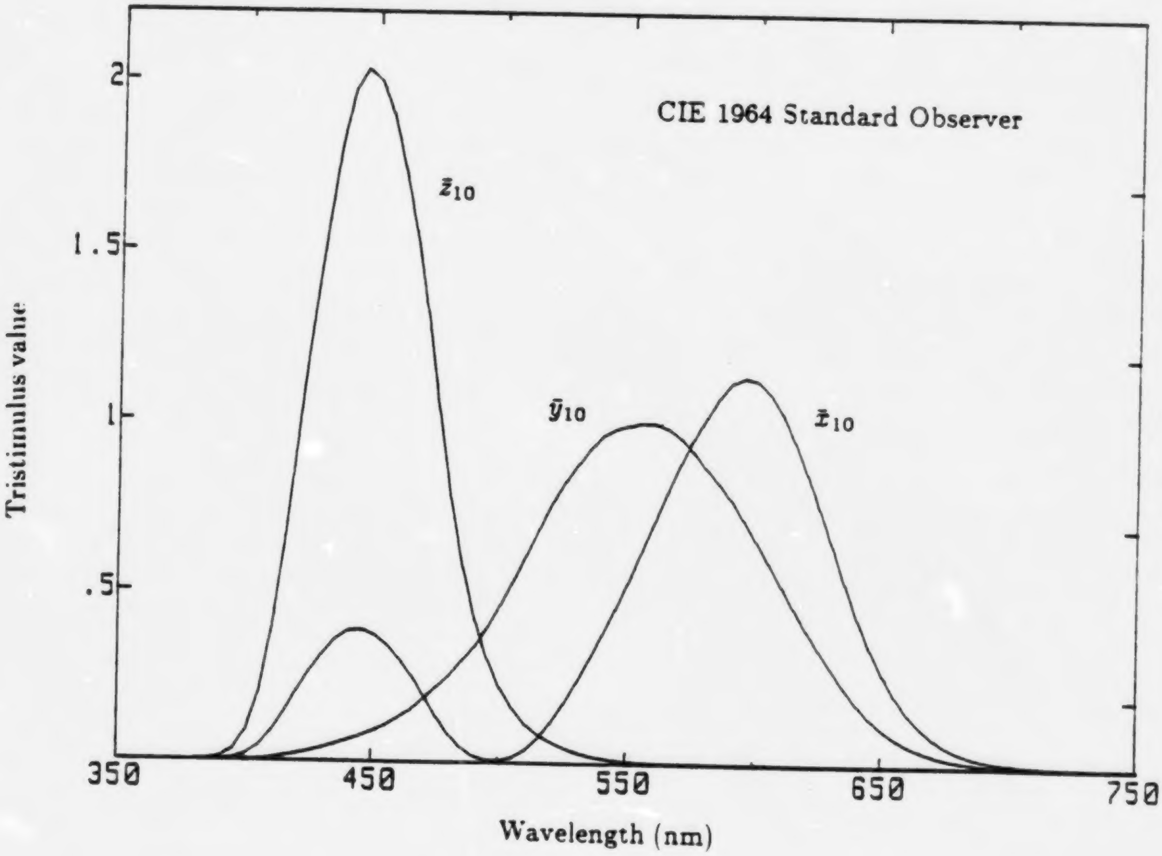

Figure 4.9. Colour-matching functions of the CIE 1964 Standard Observer.

4.7 References

Billmeyer, F.W., and Saltzman, M., 1981, *Principles of Color Technology*, New York: Wiley.

CIE*, 1983, *Publication No. 18.2, The Basis of Physical Photometry*, Paris: Commission Internationale de l'Eclairage.

CIE*, 1986 *Publication No. 15.2 Colorimetry*, Paris: Commission Internationale de l'Eclairage.

Judd, D.B. and Wyszecki, G., *1975, Color in Business Science and Industry*, New York: Wiley.

Wright, W.D., 1969, *The Measurement of Color*, London: Adam Hilger.

Wyszecki, G. and Stiles, W.S., 1982, *Color Science: Concepts and Methods, Quantitative Data and Formulae*. New York: Wiley.

---

*CIE Publications are available from the Publications Distribution Office of the National Research Council.

Notes

Lecture 5: Colin Ware

COLOUR APPEARANCE

5.1 Introduction

Before describing specific factors influencing colour appearance it is important to set the stage by having some kind of an answer to the question of what human colour vision is for. We need a general functional framework within which to set the specific processing operations which are carried out in the retina of the eye and in the brain. The way to approach this question about utility is to look at the human animal in terms of evolution and ask what is the survival value of colour vision?

In order to survive we need to be able to manipulate objects in the environment and determine their properties. Notice that information about the quality of illumination is of very little use to us. Illumination is a prerequisite for sight but otherwise we do not need to know whether the light we are seeing by is reddish because it is late in the day, or greenish because it is filtered through the canopy of leaves and branches overhead. What we do need to know about are objects - food, tools, plants, animals, other people, etc. And we can find out a lot about objects from knowledge of their surface properties, in particular, knowledge of their spectral reflectance characteristics - in everyday language - their colour. A green sheen on meat denotes that it is putrid; an orange orange, as opposed to a green orange, is ripe. Certain animals and birds can only be distinguished by their colour and some colour changes indicate whether an animal is angry or sexually receptive. All kinds of general properties of materials can be deduced from their colour, for example, green materials are usually organic. As we shall see, there is considerable evidence that what the human visual system has evolved to do is extract information about the surface properties of objects often at the expense of losing information about the quality and amount of light entering the eye. This phenomenon, that we experience coloured surfaces and not coloured light or coloured illumination, is called colour constancy. When we are talking about the apparent overall reflectance of a surface it is called lightness constancy.

The coloured environment is a very complex and intricate affair, with light bouncing in all directions being transformed at each bounce until some of it ends up in the eye of a human observer. How much of this vastly complex array of information is the human observer capable of extracting? This is a very practical problem because there is no point in modelling aspects of the environment which the observer cannot see.

The first and most useful limitation of human vision is trichromacy. This has already been dealt with at length. But notice the important consequence of trichromacy in terms of savings - because of trichromacy if we are dealing with optical mixtures of lights we need only deal with three numbers once the tristimulus values have been determined.

After trichromacy the most important property of human colour perception is level invariance. That is, the human eye does not signal the absolute level of radiation coming from a scene. Rather, within a very broad range of perhaps four orders of magnitude the eye only signals the relative amount of light coming from different parts of the environment.

5.2. Adaptation

Adaption can be roughly characterised as a gain reduction in a receptor because of an increase in the level of stimulation of that receptor. Thus adaptation means that on a bright day receptors will have a relatively low gain and on a dull day receptors will have a relatively high gain. It takes a while for this gain change to take place, especially when moving from bright to dark. Thus it can take 20 minutes or half an hour for your eyes to fully recover when you come out of sunshine and into a dimly lit room.

5.2.1 Chromatic adaptation

When illumination is coloured, different classes of cone receptor will undergo gain changes independently. Thus, if the illumination contains a lot of blue light the short wavelength cones will become less sensitive relative to the others. The effect of this is to shift the neutral point at which the three receptor types are in equilibrium such that more blue light will have to enter the eye from a surface for the surface to seem white. This is of course, exactly what is necessary for colour constancy since more blue light does enter the eye from a white surface when the illumination is blue tinted. Thus this is a way for the eye to discount the colour of the illuminant. That adaptation is both complete and effective in maintaining constancy is evidenced by the fact that not many people are aware of how much yellower ordinary tungsten room lighting is when compared to daylight. Nevertheless, although under casual observation surface colours appear to remain constant irrespective of the illumination, careful studies have suggested that subtle changes in appearance do in fact occur, (Helson et al., 1952), indicating that adaptation is not a perfect instrument of surface colour constancy. Also, adaptation does not take place entirely at the level of the receptors, post receptoral processes are also needed to explain its effects. For a review see Boynton (1979).

¿From the point of view of the TV engineer, or anyone who uses colour monitors, the differential adaptation of receptors has the happy consequence of making the eye relatively insensitive to poor colour balance. If the gain of the red gun is too high the eye will compensate by reducing the sensitive of cones most sensitive to red light. Of course since the state of adaptation of the eye is governed by the light in the entire visual field and a TV screen only takes up part of the field this adaptation will necessarily be incomplete.

5.2.2 Night vision

There is a lower limit for adaptation (as well as an upper limit where physiological damage occurs). At very low levels of illumination, e.g. starlight, cones cease to function and thus we entirely lose all sensations of chromatic colour. However, rod receptors continue functioning to give form and spatial vision (although with low acuity) as well as the lightness - darkness dimension of colour. Vision at these low levels is called 'scotopic' in contrast with 'photopic' vision which is mediated by cones.

5.3. Lateral Inhibition and Simultaneous Contrast

The second mechanism of level invariance is lateral inhibition. Because of lateral inhibition the eye only transmits edge contrast information to the brain. Lateral inhibition can be characterised in its simplest form by a connection between adjacent nerve cells such that if one cell is stimulated the adjacent cell is inhibited - becomes less sensitive (Hubel and Wiesel, 1961). There are neurons in the cortex which respond in an analogous fashion to coloured stimuli. The effect of an array of such neurons, each inhibiting its neighbour, being stimulated by a bright edge is an enhanced response just to the bright side of the edge and a depressed response just to the dark side of the edge, and an intermediate response to the uniform areas on either side. The great advantage of this mechanism is that it means that edge contrast is level invariant. That is, if you consider a black piece of coal lying on a grey pathway the contrast signal from the inhibitory net to the edge of the coal will be the same irrespective of the absolute level of illumination. Thus we can tell that an object is black by abstracting the ratio of light coming from various parts of the scene and assigning black to the region with the lowest relative level, grey to intermediate levels and white to the highest levels. Notice how absolute judgements of the light coming from an object would lead us into error - for a piece of coal on a bright day reflects more light to the eye than a piece of white paper on a dull day.

5.3.1 Edges

The inhibitory mechanism can be considered as an edge extraction device which signals the position and contrast of edges.

The perceptibility of an edge is related to the relative difference between the luminances of the abutting surfaces and not their absolute difference. Thus for surfaces having luminances L_a and L_b,

$$\text{Perceived edge} = K\frac{L_a - L_b}{L_a + L_b},$$

where K is a constant.

5.3.2 Illusion and reality

In flat displays with large areas of uniform colour, contrast causes optical illusions. See figure 5.1 for an example.

Colours may be drastically changed by other surrounding colours and uniform regions may be perceived as non-uniform. In the normal environment, as opposed to the impoverished one of man made displays, the same factors allow us to perceive the surface colours of objects with remarkable accuracy. It is as though the brain is confused in the case of the man made display and while seeking to gain meaning about light and shade causes distorted colours to be perceived.

5.3.3 Rules of contrast

Although contrast in the sense of illusion is rare in the normal environment when we do encounter it in special display situations it would be as well to know what rules it obeys. That is, how are coloured patches changed when other coloured patches are placed adjacent to them? In the achromatic domain the answer is relatively simple. That is, adding a brighter region induces darkening in an adjacent patch of colour. The amount of darkening depends on both the brightness of added region how large it is and how close it is (Jameson and Hurvich, 1972). When complex scenes are involved the perceived brightness relationships become more complex and while there is one model which attempts to account for perceived highness in random 'Mondrian' patterns (Land, 1971) there is none which even attempts to account for perceived lightness in the everyday environment.

In the chromatic dimensions the situation is more complex. The usual art book rule is that colours induce their complementaries. Thus red should induce blue-green, blue should induce yellow and green should induce purple (for complementaries defined by the CIE chromaticity diagram). However, a recent study of simultaneous colour contrast (Ware and Cowan, 1982) found the rules to be more complex. Although the departures from the complementary rule are small for most colour combinations there is one situation when that rule is strikingly in error. Since yellow is the complementary of blue the complimentary rule suggests that a blue surround on yellow should increase the yellowness (i.e. the saturation). In fact what happens is that a blue surround causes yellow to shift towards red - it becomes orange.

5.3.4 Brown and orange

If an orange patch of light is surrounded by a ring of brighter white light it will look brown. Similarly, olive green is yellow with a bright surround. In both of these instances contrast is essential in defining the perceived colour.

5.4. Shadows and Other Edges

The brain does a lot more than simply assess contrast relationships in determining colour. One of the most important operations which the visual system must do is distinguish which environmental changes in colour correspond to which of the following five types of edge:

1. Illumination edges due to shadows falling on the object.
2. Illumination edges due to the shape of the objects.
3. Colour edges at the boundary of an object of a different colour to its background.
4. Changes in the surface colour of an object (e.g. where a stripe has been painted on an automobile).
5. Colour edges at the boundary of a transparent object.

The brain has visually to extract objects and shadows from the mase of edges which is the result of early visual processing. Although research is only just beginning on how the human brain does this it is possible to sketch out some of the important variables:-

5.4.1 Fuzzy shadows

One of the most important clues we have to disambiguate edges is that shadows generally have detectably fuzzy boundaries at normal viewing distances.

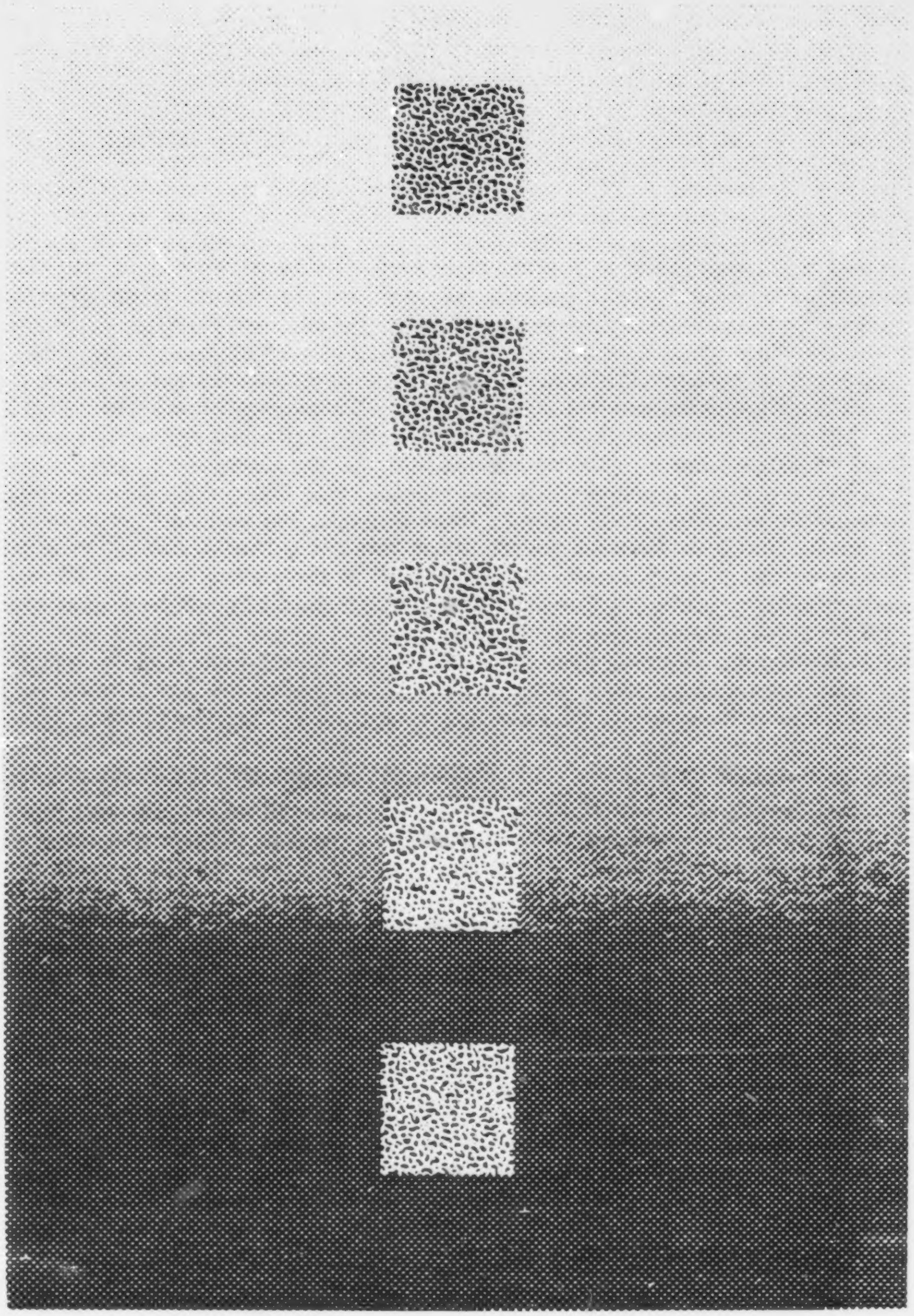

Figure 5.1. Demonstration of simultaneous lightness contrast. The small square at the top looks darker than the one at the bottom although they are actually the same. The effect is reduced somewhat by the presence of texture. It is enhanced when the viewing distance is greater and the texture becomes invisible.

5.4.2 Darkening ratios for shadows

The same shadow may fall across a number of objects. For each of these objects the luminance ratio between the area directly illuminated and the area in shadow will be roughly a constant at the shadow boundary. This constant ratio across objects makes it less likely that the lightness or colour change is surface colour change and more likely that it is in fact a shadow.

5.4.3 Continuity of shadows

The gestalt principles of form perception (Koffka, 1935) are also important in the sorting of edges. Of particular interest is the principle of good continuity. Thus if a darkening is apparent as a continuous edge crossing other boundaries this is strong evidence for a shadow.

5.4.4 Spatial arrangement

Gilchrist has shown the perception of arrangement in space to be important in simultaneous contrast. If two surfaces are perceived to be in the same plane contrast effects will be greater than if they are perceived not to be in the same plane (Gilchrist, 1979).

5.4.5 Transparency

The perception of a transparent object depends on both the right colour relations existing and the right form relations existing. If these conditions are fulfilled then what will be seen is a transparent object overlaying other objects. If they are not, what will be seen will be objects having patches of different surface colour. To show how the perception of transparency can influence the perception of colour it is necessary to do a thought experiment since the use of coloured illustrations in these notes is not feasible. Imagine a piece of yellow transparent glass overlaying a blue object. Where the glass overlaps the blue object the effect of the yellow filter on the broad band blue will be to give a light which would be seen as green in isolation. However, given the perception of transparency what we perceive is a yellow object and a blue object and no green objects. If the perception of transparency breaks down what we will see is a blue object with a green patch on it, that is, we will perceive an object which is partly green.

5.5. A Hierarchy Leading to Form Perception

What has been presented to this point is a lot of fragmented processes - adaptation, edge extraction, sorting, etc., without any overall structure. These processes are usually thought of as arranged as a sequence of processing stages. Thus, first adaptation occurs, next contrast and edge detection, next edge sorting and finally object classification. However, there is mounting evidence that the flow of information is not always in one direction only. An example of a high level process altering low level processes is that paying attention to part of a scene can cause changes in pupil diameter which will in turn affect image brightness and focus and hence all visual processes from the bottom up.

The phenomenon of memory colour gives another example of how higher processes can influence colour perception. An early study (Bruner et al., 1951) showed that when subjects were told to match the colour of certain meaningful shapes they added more colour (red for a lobster claw shape, yellow for a banana shape) than they did when asked to match the colour of purely abstract shapes. The actual colours of the shapes were always physically the same. Thus it seems that knowledge of the usual colour of an object can influence the perception of the colour of a particular example.

5.6. Miscellaneous Other Factors

In addition to the factors mentioned about there are literally hundreds of other facts about colour perception which relate to appearance but are little understood. Three of the more significant are given in the following list.

1. The texture of a surface affects its colour.
2. The gloss of a surface affects its colour.
3. The opposite phenomenon to contrast occurs in some patterns. This phenomenon is called irradiation or colour assimilation and is characterized by the colour of small display elements apparently tinting larger display elements.

5.7. Practical Considerations

The material given above is intended to give a general understanding of some of the broader issues involved in colour appearance. What follows is a list of some practical concerns relating to appearance.

5.7.1 Colour scales for maps

It is common practice to use chromatic scales or grey scales for indicating height on geographical maps or energy levels on astronomical maps. Colour scaling makes the shape of the surface immediately visible. However, if absolute height values are to be read using a key large errors can occur because of simultaneous contrast effects. This is especially apparent with grey scale maps.

5.7.2 Colour differences

Colour difference formulae (discussed in lecture 15) may give the impression that it is possible to create colour sets for which the perceived differences are known. This may be possible for simple environments where the illumination is standardised, but in environments where the illumination and shapes are complex, contrast effects will cause major distortions in the perceived colour differences.

5.7.3 Surround lighting

When concerned about the colours in a display it is important to consider the surround lighting. Since the state of adaptation of the eye is determined by the entire visual field, not just what is being attended, the lighting in the periphery of vision is critical.

5.7.4 Absolute values

Since the eye is very insensitive to absolute values, especially luminance values careful colorimetry designed to accurately measure or specify absolute values is usually a waste of time.

5.7.5 Relative values

The eye is very sensitive to relative values, particularly at abrupt boundaries and so accurate relative colorimetry is important. Great attention should be paid to contrast relationships when working with coloured displays.

5.7.6 Brightness Perceived brightness and luminance are not related

The perceived brightness of display lights is relative to the background on which they are displayed thus perceived brightness and relative luminance are related. Nevertheless if the display is complex then the contrast interactions are likely to be sufficiently complex that it is not possible to predict perceived brightnesses.

5.7.7 Heterochromatic brightness

It is an empirical phenomenon that even it they have the same background, lights of the same luminance but different colours are perceived as having different brightnesses. For this purpose a conversion formula has been developed for calculating the luminances of stimuli which are equal in brightness (Ware and Cowan, 1983). The following brief description describes how to use the conversion factor.

The conversion function may reliably be used under the following conditions:

1. The stimuli subtend 30 minutes of visual angle, or more.
2. The stimuli have luminance 2 candela per square meter, or greater.
3. The stimuli to be compared have similar surround conditions.
4. The ambient light is relatively neutral.

To compare two stimuli:

1. Determine the X, Y, Z tristimulus values of each stimulus, using the CIE 1931 color matching functions.
2. Calculate the (x, y) chromaticity coordinates and luminance (L_0) of each stimulus.
3. If the y chromaticity coordinate of a stimulus is greater than 0.02, calculate its conversion factor using the equation

$$C = 0.256 - 0.184y - 2.527xy + 4.656x^3y + 4.657xy^4$$

4. **Otherwise, if the y chromaticity coordinate is less than 0.02, determine the conversion factor from the dominant wavelength of the stimulus, using the graph in Ware and Cowan (1983).**
5. **For each stimulus, calculate**

$$\log(L) + C$$

or alternatively

$$L \cdot 10^C$$

Whichever stimulus has the greater value on either of these measures will be seen as brighter by the average observer. If the two stimuli have equal values, they will be seen to be equally bright.

5.8. References

Boynton, R. M., 1979, *Human Colour Vision*, New York: Holt, Rinehart and Winston.

Bruner, T. S., Postman, L. and Rodrigues, J., 1951, 'Expectation and the Perception of Color', *American Journal of Psychology*, **64:** 216-227.

Gilchrist, A. L., 1979, 'The Perception of Surface Blacks and Whites', *Scientific American*, **238:** (March), 88.

Helson, H., Judd, D. B. and Warren, M. H., 1952, 'Object-Color Changes from Daylight to Incandescent Filament Illumination', *Illumination Engineering*, **47:** 221.

Hubel, D. H. and Wiesel, T. N., 1961, 'Integrated Action in the Cat's Lateral Geniculate Body', *Journal of Physiology*, **155:** 385-398.

Jameson, D. and Hurvich, L. M., 1972, 'Color Adaptation: Sensitivity, Contrast, After-Images', in *Handbook of Sensory Physiology*, **VII/4:** 568-581, D. Jameson and L. M. Hurvich, eds., New York: Springer Verlag.

Koffka, K., 1935, *Principles of Gestalt Psychology*, New York: Harcourt and Brace.

Land, E. H., 1971, 'The retinex theory of color vision', *Scientific American*, 237, 108-128.

Ware, C. and Cowan, W. B., 1982, 'Changes in Perceived Color Due to Chromatic Interactions', *Vision Research*, **22:** 1353-1362.

Ware, C. and Cowan, W. B., 1985, 'Specification of Heterochromatic Brightness Matches: A Conversion Factor for Calculating Luminances of Stimuli Which are Equal in Brightness', National Research Council of Canada Technical Report, Number 26055.

Notes

Lecture 6: A. R. Robertson

COLOUR ORDER SYSTEMS

6.1 Introduction

A colour order system is a systematic method of sampling the surface colour solid. Most systems are illustrated by a collection of samples known as a colour atlas. In a colour atlas a reasonable number of samples (typically about 1000) is selected to cover the range of all possible colours and the samples are arranged on pages according to the chosen system.

Reviews of the principles of colour order systems are given by Billmeyer and Saltzman (1981), Judd and Wyszecki (1975) and Wyszecki and Stiles (1982).

6.2 Types of Colour Order Systems

Three types of colour order systems have been distinguished.

6.2.1 Colour Mixture Systems

In these systems, the colours produced by systematic variation of the primaries in an additive mixture system (such as a rotating-sector disk) are duplicated by material samples.

6.2.2 Colorant Mixture Systems

These systems are based on the regular adjustment of a limited number of dyes or pigments. Their main purpose is to illustrate the gamut and other properties of a particular set of colorants and, in some cases, to enable recipes for intermediate colours to be derived.

6.2.3 Colour Appearance Systems

The scales of these systems are chosen to represent attributes of perceived colours. In many ways they are the most important of the three types because they are not linked to any particular technology. The rest of this lecture will concentrate mainly on colour appearance systems although some of the principles apply to the other types too.

6.3 Dimensions of Colour Order Systems

Object colours, viewed in fixed conditions of observation, can always be described in terms of three independent component attributes or dimensions. However, the choice of which three attributes to use varies significantly from one system to another. A comparison of the attributes of various systems has been made by Robertson (1984).

6.3.1 Hue, Lightness and Chroma

Perhaps the most obvious way in which colours differ from each other is described by the attribute hue. Hue is the attribute according to which a colour appears to be similar to one or to proportions of two, of the perceived colours red, orange, yellow, green, blue, and purple. As this is a closed scale (the next step after purple is red again), the most appropriate geometrical locus to represent hue is a circle.

Within each hue, there are still a large number of different colours. To many people the next most obvious attribute is lightness. This is the attribute according to which a sample appears to reflect a larger or smaller fraction of the incident light.

The quantity frequently chosen as the third attribute (to describe the remaining variation among object colours of equal hue and lightness) is called chroma. This is the degree to which a perceived colour differs from an achromatic perceived colour of the same lightness.

Figure 6.1 shows the geometrical model most appropriate for representing the colour solid using the dimensions of hue, lightness and chroma. The achromatic (grey) perceptions range from black to white and are represented by points on the vertical axis, the lightness axis. All chromatic perceptions are represented by points away from the vertical axis. Colours of constant lightness are represented by points in horizontal planes. Colours of constant hue are represented by points in vertical planes radiating from the central vertical axis, with the hue being measured by the angle from a reference plane. Colours of constant chroma are represented by coaxial cylinders with the radius being the measure of chroma.

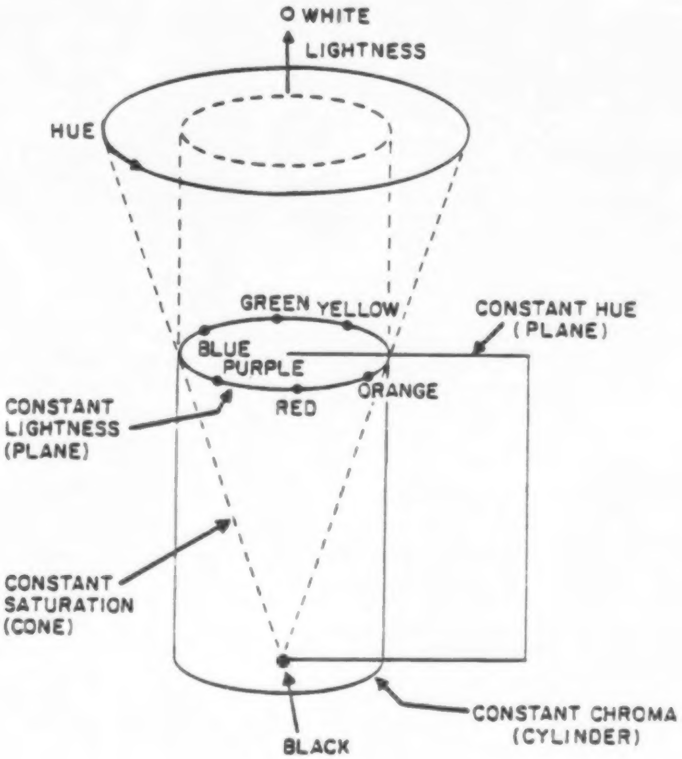
WHITE
LIGHTNESS
HUE
GREEN
YELLOW
CONSTANT HUE (PLANE)
BLUE
PURPLE
ORANGE
RED
CONSTANT LIGHTNESS (PLANE)
CONSTANT SATURATION (CONE)
BLACK
CONSTANT CHROMA (CYLINDER)

Figure 6.1. Geometrical model of colour space based on the attributes of hue, lightness and chroma.

A variant of chroma is saturation. This is the proportion of chromatic colour in the total sensation (as opposed to chroma which can be thought of as the amount of chromatic colour). In the geometrical model of figure 6.1, perceived colours of constant saturation lie on a cone whose apex is at the black point. The importance of saturation lies in the so-called 'shadow series' of colours. As an object becomes progressively more shaded from a light source both its perceived lightness and chroma decrease, but its saturation remains constant.

6.3.2 Resemblances to elementary colours

An entirely different way to describe a perceived colour is in terms of its degree of resemblance to six elementary colours red, green, yellow, blue, black and white. These elementary colours are theoretical ones in which no trace of the others can be seen.

In this system the hue of a colour is defined in terms of the resemblance to the two nearest chromatic elementary colours. Thus an orange might be said to bear a 30% resemblance to red and a 70% resemblance to yellow. This definition of hue is very similar to the definition in the hue, lightness, chroma system.

The second attribute used in this system is called chromaticness. This is the degree of resemblance to the colour of the same hue having the maximum possible chromatic content. The third variable can be either blackness or whiteness, the degree of resemblance to the elementary black or elementary white respectively.

A geometrical model that can be used for this system is shown in figure 6.2. The elementary colours are represented at the ends of three mutually perpendicular axes, green-red, blue-yellow, and black-white. Colours of constant hue again lie on vertical planes. Colours of constant chromaticness lie on cylinders, colours of constant blackness lie on cones pointing upwards, and colours of constant whiteness lie on cones pointing downwards.

6.3.3 Regular rhombohedral lattice sampling

Another possible criterion for a colour order system is the completely uniform spacing of colours. Polar coordinate systems are not useful in this regard because the interior of the colour solid is sampled more densely than the outer regions. The geometrical model that is most convenient is the regular rhombohedral lattice shown in figure 6.3. Each point in the lattice is surrounded by twelve other points all equidistant from it.

It is possible to sample colour space using such a lattice so that each colour differs from its nearest neighbours by an equal amount. This does not necessarily allow the axes of the system to represent any particular attribute.

6.3.4 Colour naming

A fourth principle for a colour order system is that of colour names. It is possible to choose a limited series of simple names such as red, green, brown, olive, pink etc. and to combine them with adjectives such as light, dark, strong, vivid etc. to produce an elementary colour order system. The number of colours that can be included in such a system is limited by the number of meaningful names that can be chosen.

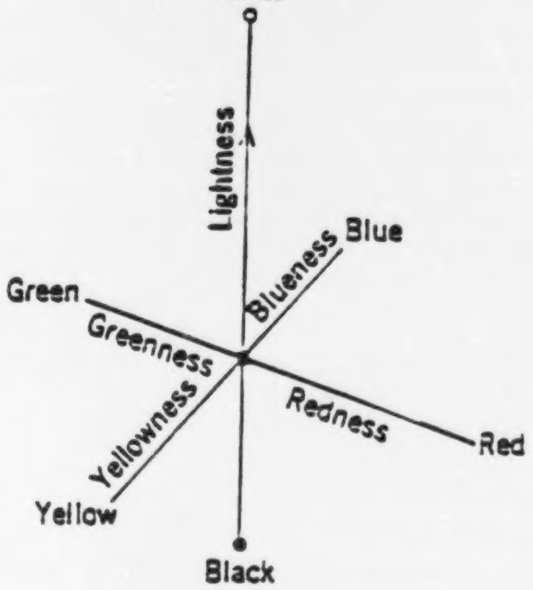
White
Lightness
Blue
Blueness
Green
Greenness
Yellowness
Redness
Red
Yellow
Black

Figure 6.2. Geometrical model of a colour space based on resemblances to six elementary colours. (**Judd and Wyszecki, 1975**).

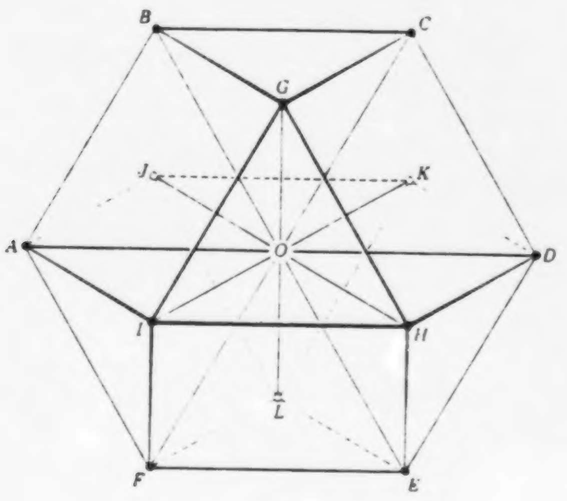

Figure 6.3. Regular rhombohedral lattice showing the 12 nearest neighbours of the point. O (Judd and Wyszecki, 1975).

6.4 Examples of Colour Order Systems

6.4.1 Munsell System

The colour order system used most commonly in North America is the Munsell System. This is a colour appearance system based on the hue-lightness-chroma principle.

The hue scale is divided into 10 segments labelled Y, GY, G, BG, B, PB, P, RP, R and YR. The letters indicate approximate correlations with the names yellow, green, blue, purple and red. The scale can be further divided by a numerical prefix in the range 0 to 10 (0 GY is the same as 10 Y etc. with the latter notation being preferred).

The lightness scale, called value by Munsell, ranges from 0 (black) to 10 (white). Munsell chroma is 0 for achromatic colours and increases to a maximum that depends on the hue and on the particular colorants used to produce the system.

The Munsell notation is always given in the order hue, value, chroma according to the following form:

5RP 6/8

The organization of colours of constant hue and of constant value in the Munsell Book of Color is illustrated in figures 6.4 and 6.5.

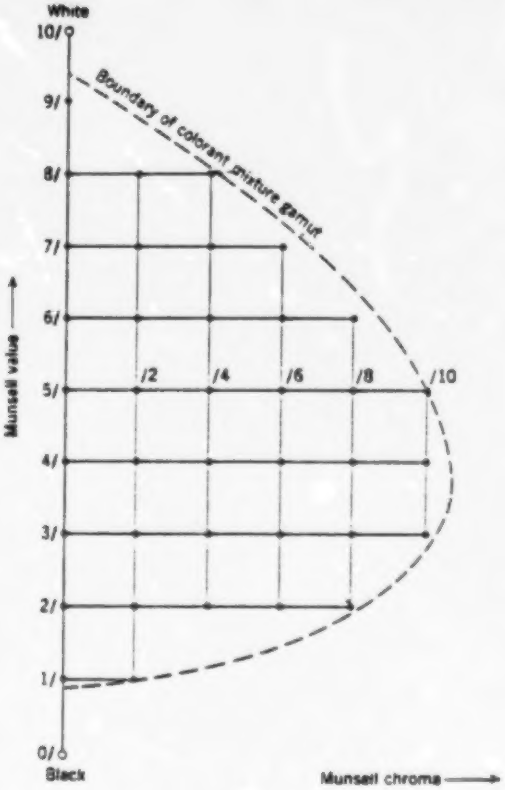

Figure 6.4. Organization of colours of constant Munsell hue (Judd and Wyszecki, 1975).

Each of the three scales of the Munsell System is arranged so that equal steps represent equally perceptible differences of the attribute considered, when the other two attributes are held constant.

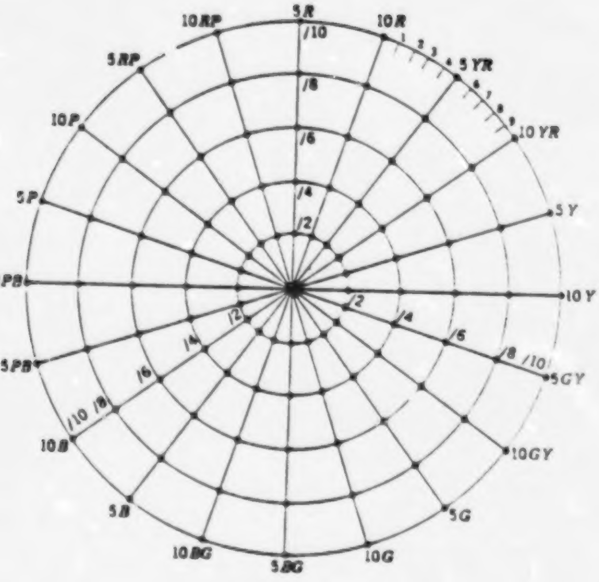

Figure 6.5. Organization of colours of constant Munsell value (Judd and Wyszecki, 1975).

6.4.2 Natural Colour System (NCS)

This is a Swedish system which is based on the resemblance principle.

The hue notation is illustrated in figure 6.6.

In a constant hue plane, the samples of the NCS atlas are in a triangular arrangement as shown in figure 6.7. Horizontal lines do not represent planes of constant lightness however because each of the purely chromatic colours (C) has a different lightness.

6.4.3 DIN Colour System

This is a German system based on the hue-lightness-saturation principle.

6.4.4 Ostwald System

This system is often referred to in the literature although atlases illustrating it are no longer available. It is a colour-mixture system with the samples arranged in a way similar to those of the NCS.

6.4.5 Optical Society of America System

At present this is the only example of a system based on the regular rhombohedral lattice principle. It contains about 500 colours, each equally distant from 12 nearest neighbours.

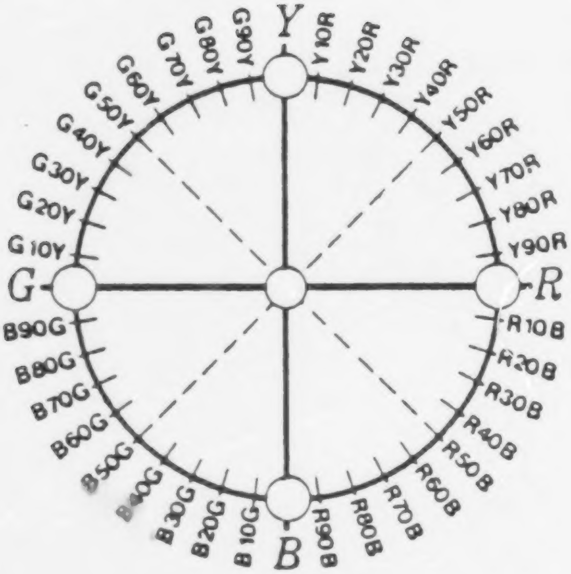
Y
Y10R
Y20R
Y30R
Y40R
Y50R
Y60R
Y70R
Y80R
Y90R
R
R10B
R20B
R30B
R40B
R50B
R60B
R70B
R80B
R90B
B
B10G
B20G
B30G
B40G
B50G
B60G
B70G
B80G
B90G
G
G10Y
G20Y
G30Y
G40Y
G50Y
G60Y
G70Y
G80Y
G90Y

Figure 6.6. Hue circle of the NCS (Hård and Sivik, 1981) It is important to note that the elementary hues are deliberately placed 90° apart . Thus the spacing (in terms of perceived colour differences) is different in each quadrant.

6.4.6 ISCC-NBS System

The Inter-Society Color Council - National Bureau of Standards system is a color-naming system. A total of 267 colours is included using the hue names and modifiers given in table 6.1, together with the names white, grey and black.

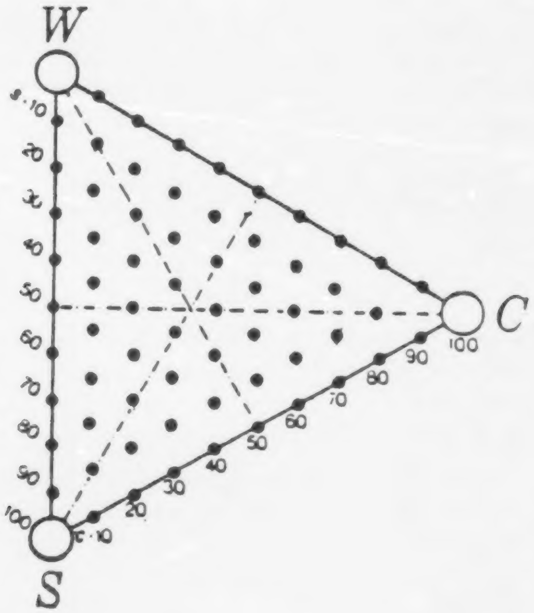
W
C
S
s·10
20
30
40
50
60
70
80
90
100
c·10
20
30
40
50
60
70
80
90
100

Figure 6.7. Constant hue plane of the NCS (Hård and Sivik, 1981)

Hue Names	Modifiers
Red	Very pale
Orange	Pale
Yellow	Light
Green	Moderate
Blue	Dark
Violet	Very dark
Purple	Brilliant
Pink	Strong
Brown	Deep
Olive	Very deep
	Vivid

Table 6.1. Hue names and modifiers in the ISCC-NBS method of designating colors.

6.5 Relation to CIE System

The samples of the colour atlases illustrating almost every colour order system have been measured in CIE terms so that intercomparisons can be made. However such intercomparisons are usually tedious and difficult because the data are in tabular form and not represented by mathematical formulae.

6.6 Conclusion

Many colour order systems are available. The principles on which they are based are often different and sometimes incompatible. The best system to use depends to a large extent on the purpose for which it will be used.

6.7 References

Billmeyer, F.W., Jr. and Saltzman, M., 1981, *Principles of Color Technology*, New York: Wiley.

Hård, A. and Sivik, L., 1981, 'NCS - Natural Color System: A Swedish standard for color notation', *Color Res. Appl.* **6: 129-138**.

Judd, D.B. and Wyszecki, G., 1975, *Color in Business, Science and Industry*, New York: Wiley.

Robertson, A.R., 1984, 'Colour Order Systems - An Introductory Review', *Color Res. Appl.* **9**: 234-240.

Wyszecki, G. and Stiles, W.S., 1982, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, New York: Wiley.

Notes

Notes

Lecture 7: A. A. Gaertner

PRINCIPLES OF PHOTOMETRY AND RADIOMETRY

7.1 Introduction

The next series of lectures will be concerned with some of the practical problems we encounter when actual measurements must be made. This lecture will introduce you to some of the jargon for which the field of light measurements is infamous. We will primarily consider various geometrical arrangements of the light source and detector which have become, because of their extensive application, basic measurement configurations.

7.2 Basic Light Measurement

Whenever you require the colorimetric properties of a source, the simplest system you will have to set up is a basic measurement system such as the one shown in figure 7.1. The source can be either a lamp of some type, or an illuminated object. The detector can be as simple as a silicon diode or a photomultiplier tube. Although a source and a detector are the most basic components, in general you will find it necessary to provide additional components in the beam path which alter the character of the beam to enable you to obtain the particular information you require. For simplicity, let us call these components an analyzer.

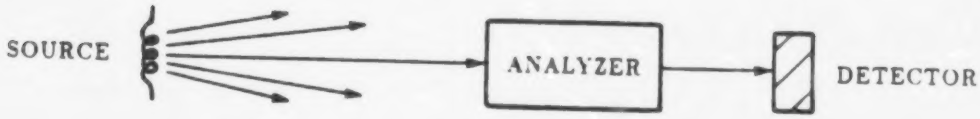

Figure 7.1. Basic measurement system

In performing any measurement, you must ask yourself two basic questions: What do I want to measure? and, What am I actually measuring?

7.3 Factors Which Affect Our Measurement

After you set up your measurement systems you will find that there are various factors or effects which will influence the results which you obtain. These are summarised in table 7.1.

1.	Position	Geometrical effects
2.	Direction	Geometrical effects
3.	Wavelength	Spectral effects
4.	Time	
5.	Polarisation	
6.	Coherence	Interference effects

Table 7.1. Measurement Parameters.

7.4 Basic Photometric and Radiometric Quantities

This lecture is concerned with the basic geometrical configurations of light measurements, their definitions, their names, and the units in which they are measured. In addition, we will have to take into consideration two other measurement parameters:

i. Spectral effects: Is the measurement photometric or radiometric?

ii. Time effects: Is the light continuous or flashed? The terms we will consider were developed for continuous light sources. Therefore the units will involve power, or flux.

The five most important quantities are presented in table 7.2.

Name	Symbol	Radiometric Unit	Photometric Unit
Irradiance	E_e	watt per square metre ($w \cdot m^{-2}$)	
Illuminance	E_v		lumen per square metre ($lm \cdot m^{-2}$) lux (lx)
Radiant Flux	Φ_e	watt (w)	
Luminous Flux	Φ_v		lumen (lm)
Radiant Intensity	I_e	watt per steradian ($w \cdot sr^{-1}$)	
Luminous Intensity	I_v		candela (cd)
Radiance	L_e	watt per steradian per square metre ($w \cdot sr^{-1} \cdot m^{-2}$)	
Luminance	L_v		candela per square metre ($cd \cdot m^{-2}$)
Radiant Exitance	M_e	watt per square metre ($w \cdot m^{-2}$)	
Luminous Exitance	M_v		lumen per square metre ($lm \cdot m^{-2}$)

Table 7.2. Names, symbols, and units of the most important radiometric and photometric quantities.

We have made a distinction between a quantity, its name, and the unit in which it is measured. The quantity is the particular geometrical arrangement of source, analyzer, and detector which we are using. Over the years, certain geometrical arrangements have become so common and useful that we started to associate names with them (quantity name). Finally, there is a unit associated with each quantity, in the same manner as the unit distance, the metre, is associated with distance measurement.

It is possible to separate the five quantities into two groups: one of which considers the light from the receiver's point of view (the quantity irradiance/illuminance), and the other 4 quantities which consider the light more from the source's point of view.

We can now look at each of these five quantities in more detail. The definitions presented are rewordings of the exact definitions which can be found in CIE (1970). Further information can be found in the other 3 references at the end of this lecture.

7.4.1 Irradiance/Illuminance (E)

This is the ratio of the radiant (or luminous) power incident on a surface, divided by the area of that surface. (figure 7.2)

$$E = \Phi/A$$

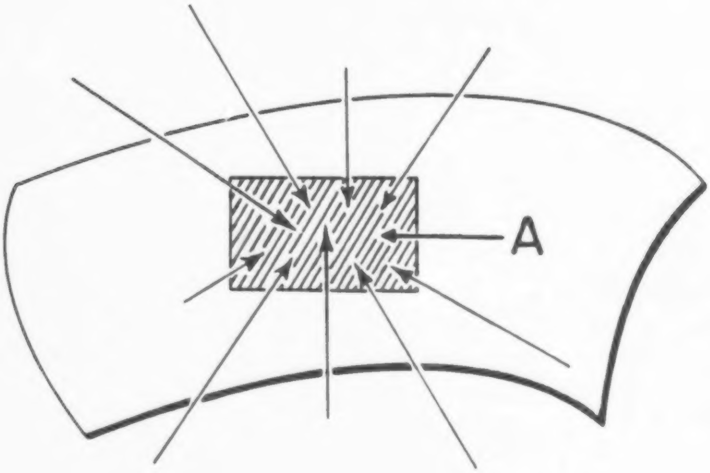

Figure 7.2. Geometrical configuration for the quantities irradiance and illuminance.

7.4.2 Flux (Φ)

Radiant flux

Power emitted, transferred, or received in the form of radiation. (Note: Radiant flux is sometimes called radiant power and denoted by the symbol P.)

Luminous flux

Quantity derived from radiant flux by evaluating the radiation according to its action upon a selective receptor, the spectral responsivity of which is defined by one of the standard spectral luminous efficiencies, $V(\lambda)$ or $V'(\lambda)$.

The total flux output of a lamp is evaluated using an instrument called an integrating sphere as shown in figure 7.3.

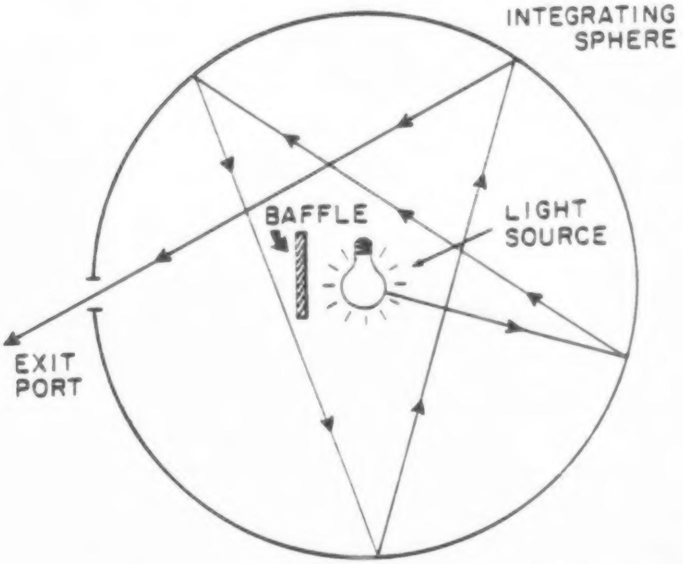

Figure 7.3. The integrating sphere used for the evaluation of the quantity flux.

7.4.3 Intensity (I)

Intensity is the ratio of the flux leaving a source and propagating in an element of solid angle containing a given direction, divided by the size of that element of solid angle. The basic principles are shown in figure 7.4.

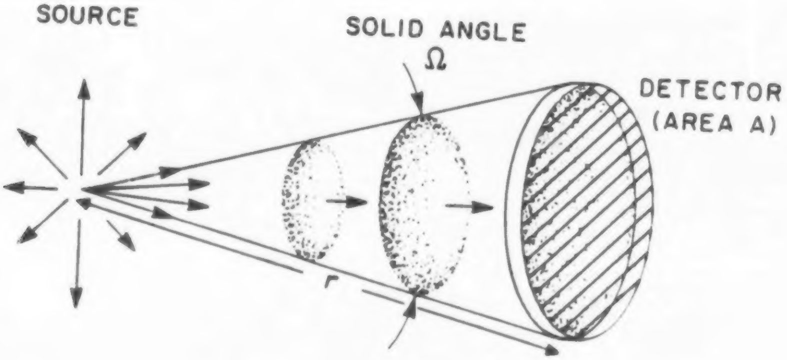

Figure 7.4. Defining principles for the quantity intensity.

$$I = \Phi/\Omega = \frac{\text{flux received by detector}}{\text{soldi angle subtended by detector}}$$

The solid angle is defined as

$$\Omega = \text{area of detector}/r^2$$

Its units are steradians.

The candela is the defining or basic unit for photometry.

7.4.4 Radiance/Luminance (L)

This is the ratio of the flux leaving a source of given area and propagating in an element of solid angle containing a given direction, divided by the size of that element of solid angle and the size of the given area of the source when projected along the given direction. The basic ideas are shown in figure 7.5.

The radiance is given by

$$L = \Phi/(A_p\, d\omega)$$

The term A_p is the projected area

$$A_p = A \cos \varepsilon$$

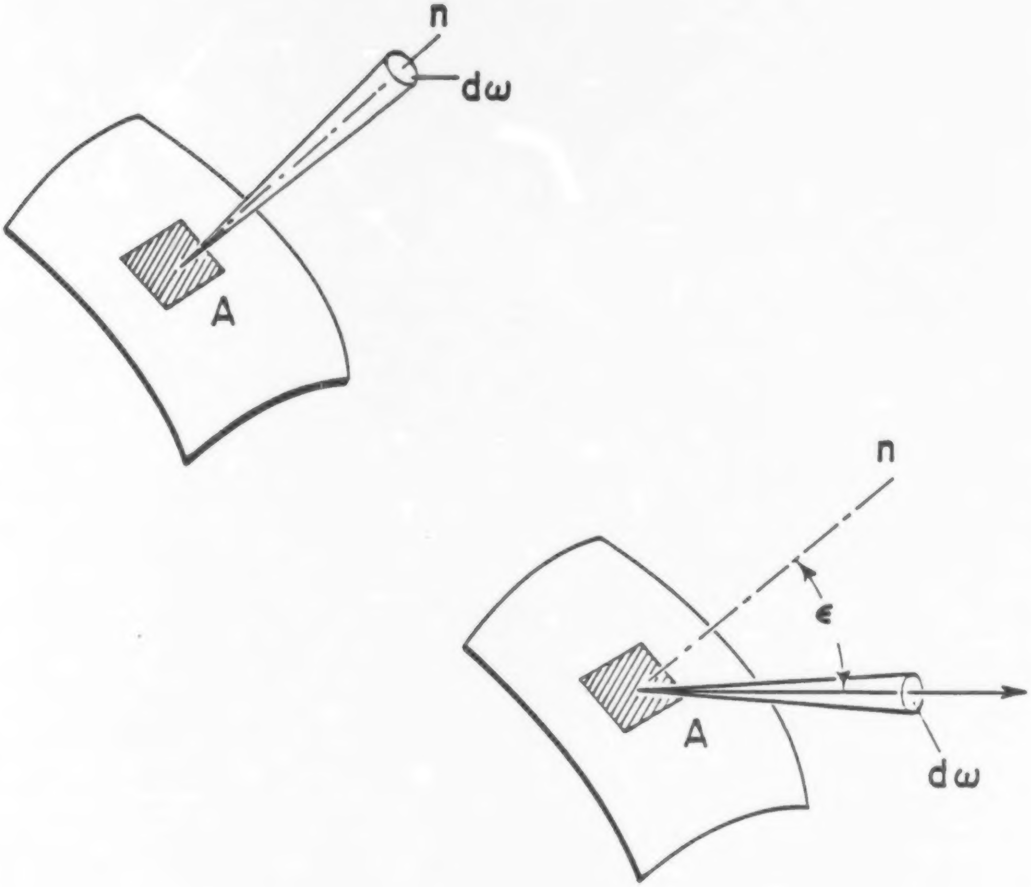

Figure 7.5. Geometrical configurations for the quantities radiance and luminance.

7.4.5 Exitance (M)

Exitance is the ratio of the total flux leaving an element of a surface, divided by the area of that surface element (figure 7.6).

$$M = \Phi / A$$

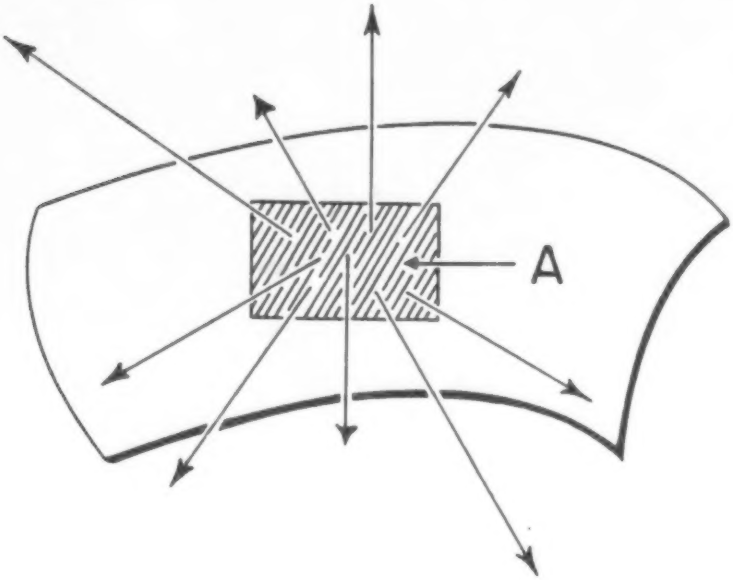

Figure 7.6. Geometrical configuration for the quantity exitance.

7.5 Applications

There are two important applications of the basic geometrical configurations which we have considered.

7.5.1 Uniformly diffusing surface (Lambert surface)

The luminance (or radiance) of this surface is the same in all directions. (figure 7.7)

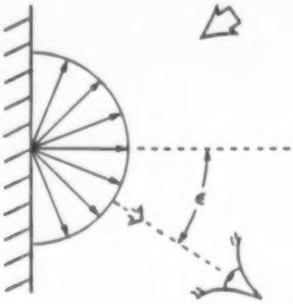

Figure 7.7. The radiance of a uniformly diffusing surface is equal in all directions.

Consequences

1. This surface will appear equally bright when viewed from all directions.
2. The intensity per unit area of this source varies as the cosine of the angle at which it is viewed. (Lambert's cosine law).

$$L = \Phi/A_p\Omega = \Phi/(\Omega A \cos\epsilon)$$

$$\Phi = LA\Omega\cos\epsilon$$

$$I = \Phi/\Omega = LA\cos\epsilon$$

$$I/A = L\cos\epsilon$$

3. This surface can be considered as a luminous or a radiant source. If the reflectance of the surface is known, its luminance or radiance can be calculated from the irradiance it receives from a source of known irradiance or intensity. Since the luminance is equal in all directions, the total power output of the surface can be integrated over all the directions to obtain the exitance of the surface:

$$M = \pi L.$$

Since the exitance is related to the illuminance through the reflectance,

$$M = \rho E,$$

we can obtain the luminance of the surface from the illuminance:

$$L = M/\pi = \rho E/\pi.$$

7.5.2 The $1/r^2$ law

This law relates the illuminance of a surface due to light emitted from a source of intensity I to the distance r between the source and the surface. The basic concepts are shown in figure 7.8.

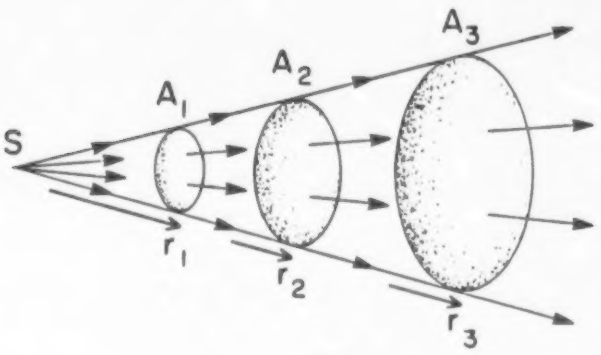

Figure 7.8. The origins of the inverse square law.

$$I = \Phi/\Omega; \qquad \Phi = I\Omega.$$

The same flux passes through each area A because the solid angle is the same for all of them:

$$\Omega = A_1/r_1^2 = A_2/r_2^2 = A_3/r_1^3$$

Thus,

$$E = \Phi/A = I\Omega/A = I/r^2.$$

Limitations

The distance r between the source S and the detector must be large enough that S can be considered a point source. A 'rule of thumb' is that if r is at least ten times larger than the largest dimension of the source or detector, the error will be less than 1%.

7.6 References

CIE*, 1970, *Publication No. 17 (E-1.1.), International Lighting Vocabulary,* 3rd Edition, Commission Internationale de l'Eclairage: Paris.

Grum, F., and Becherer, Richard J., 1979, *Optical Radiation Measurements, Volume 1, Radiometry,* Academic Press: New York.

Walsh, J.W.T., 1958, *Photometry,* Constable and Company Ltd.: London.

Wyszecki, G., and Stiles, W.S., 1982, *Color Science,* 2nd Edition, Wiley: New York.

---

*CIE Publications are available from the Publications Distribution Office of the National Research Council.

Notes

Lecture 8: A. A. Gaertner

SPECTRORADIOMETRY

8.1 Spectral Distribution Curves

The amount of electromagnetic radiation present at the different wavelengths of the visible spectrum is very important for colorimetric purposes. The data is usually presented in curves such as that shown in figure 8.1.

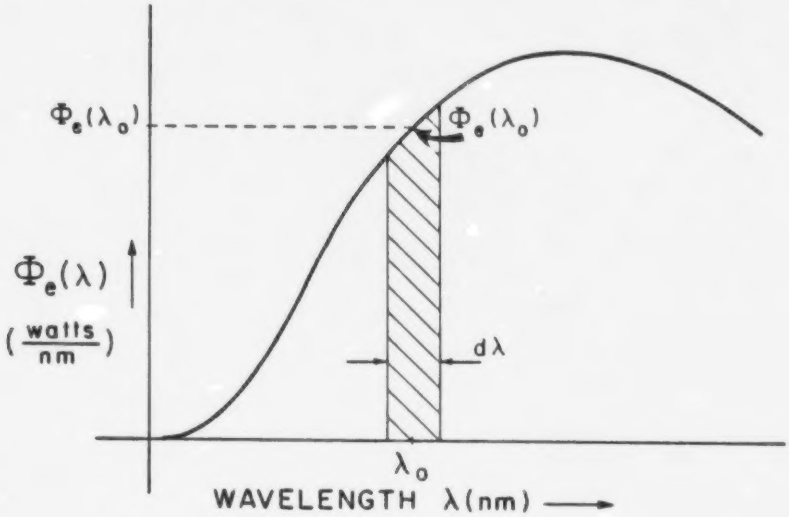
$\Phi_e(\lambda_o)$
$\Phi_e(\lambda_o)$
$\Phi_e(\lambda)$
$\left(\frac{\text{watts}}{\text{nm}}\right)$
$d\lambda$
λ_o
WAVELENGTH λ(nm) ⟶

Figure 8.1. Spectral power distribution curve.

$\Phi_e(\lambda_o)$ is the radiant flux per unit wavelength at the wavelength λ_o. Its unit is watts/nm. It is the radiant flux available to you in a 1.0 nm bandwidth at λ_o.

Therefore the flux available to you at wavelength λ_o is given by:

Flux = area of ▨ $= \Phi_e(\lambda_o)\, d\lambda$,

where $d\lambda$ is the wavelength bandwidth in which you want the flux.

8.2 Basic Spectral Measurement System

The basic components required to obtain a curve such as that shown in figure 8.1, are shown in figure 8.2. This is very similar to that shown in figure 7.1. In the present case, the main purpose of the analyzer is to obtain the spectral content of the light entering it. The combination of the spectral analyzer and detector is called a spectroradiometer.

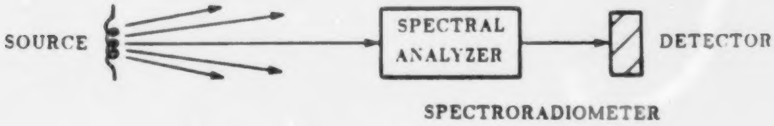
SOURCE
SPECTRAL ANALYZER
DETECTOR
SPECTRORADIOMETER

Figure 8.2. The basic components of a spectral measurement.

8.3 Spectral Analysis Methods

8.3.1 Filters

The basic ideas are presented in figure 8.3. The filter allows only a certain range of wavelengths to pass through it to the detector. The other wavelengths are either absorbed or reflected by the filter.

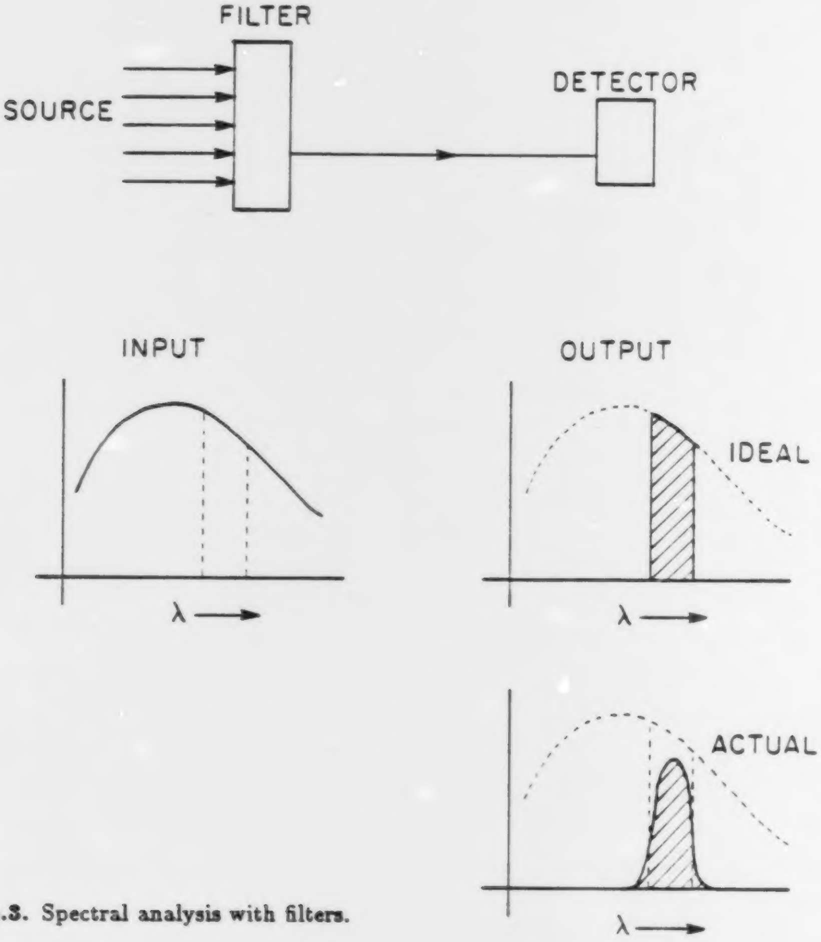
FILTER
SOURCE
DETECTOR
INPUT
OUTPUT
IDEAL
ACTUAL
λ ⟶
λ ⟶
λ ⟶

Figure 8.3. Spectral analysis with filters.

8.3.2 Monochromators

The basic components of a monochromator system for dispersing light are shown in figure 8.4.

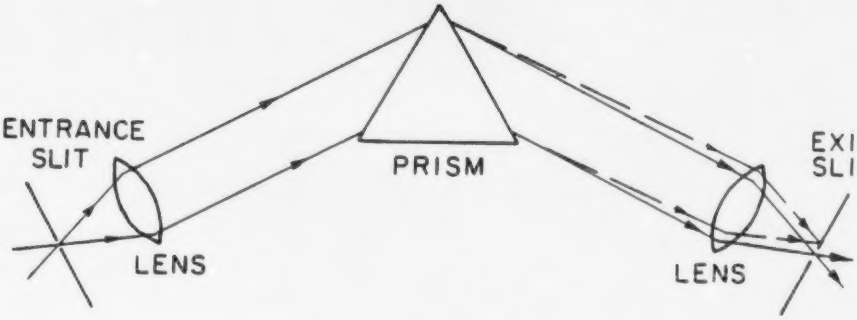

Figure 8.4. Basic components of a monochromator.

The prism (or grating) is the spectral dispersing element which takes the collimated incoming light and separates the different wavelengths by causing them to propagate in different directions once they leave the disperser. Thus the different wavelengths are physically separated.

The input lens (or mirror) serves to collimate the incoming light onto the disperser, and the exit lens focuses the dispersed light onto the exit focal plane. The entrance slit defines the portion of the incoming light which will be analyzed, and the exit slit defines the wavelength region which we will detect.

There are several design considerations which need to be taken into account when you obtain a monochromator system.

Wavelength range

The wavelength range available to you with a prism will depend on the material of which the prism is composed. The transmission and dispersion of the material are the prime factors. In the case of a grating, the factors which are important are the ruling (number of grooves per cm) and the blaze (shape of the grooves).

Throughput

The throughput is related to the maximum amount of light you can cause to pass through the monochromator. The important elements are shown in figure 8.5.

Throughput is defined as the product of the area of the slit (A) and the solid angle at the slit (Ω). It can be related to the F-number of the system which is defined as the ratio of the focal length of the lens (f) to the diameter of the lens (l). The smaller the F number, the larger the throughput.

Resolution

The resolution of a system is concerned with our ability to distinguish different features in the spectral distribution curve which are separated by small wavelength differences. The greater the dispersion of the prism material, or the finer the rulings on the grating, the greater will be the resolution. The size of the dispersing element and focal length of the lenses will also affect the resolution.

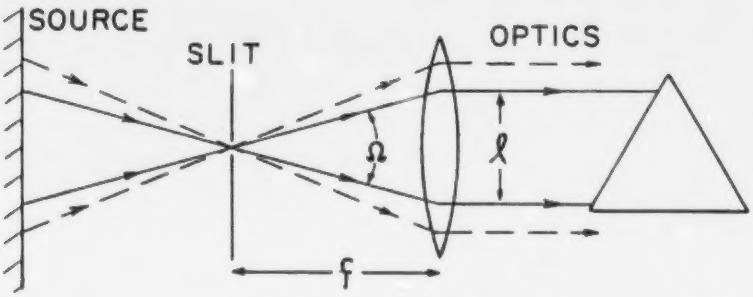

Figure 8.5. Basic concepts for the definition of throughput.

Bandwidth

The bandwidth of a system is related to the range of wavelengths and the amount of light at each wavelength in this range (or band) which is passed by the system through its exit slit.

For each wavelength in the incident beam of light, an image of the input slits is formed at the output focal plane. Since this image has a real width, there will be an overlap of the images of neighbouring wavelengths at the focal plane. Thus the spectral content of the light passed by the exit slit will depend on the size of both the entrance and exit slits as well as the dispersion of the system.

Prism vs Gratings

Some differences between systems built with a prism or a grating as the disperser are shown in Table 8.1.

	Prism	Grating
Dispersion	non-linear, low	linear, high
Order problems	no	yes
Wavelength range	materials available	grating ruling

Table 8.1. Comparison of prism and grating monochromators.

Quality of construction

The care with which the instrument has been made will show in such properties as stray light, wavelength accuracy, and wavelength reproducibility.

8.3.3 Fourier-Transform Techniques

In this technique, the wavelengths of which the light is composed are caused to interfere by modifying the phase of part of the beam. The resulting interference pattern is mathematically analysed to produce a measurement of the spectral content of the light. These techniques are predominantly used in the infrared.

8.4 Spectroradiometer System Analysis

The system shown in figure 8.6 can be used to study how a spectroradiometer modifies the light which passes through it. The effect of each component is shown in the spectral curves of figure 8.7.

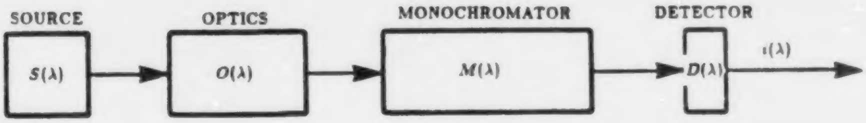

Figure 8.6. Components of a spectroradiometric measurement: $S(\lambda)$ is the spectral output of our source; $O(\lambda)$ is the spectral transmission of any input optics; $M(\lambda)$ is the spectral transmission of the monochromator; $D(\lambda)$ is the spectral responsivity of the detector.

Reasonable qualitative assumptions can be made concerning the quantities $O(\lambda)$ and $D(\lambda)$. $M(\lambda)$ is very difficult to obtain. As a result, we can determine $M(\lambda)$ from the ratio of the system response $i(\lambda)$ to the other components:

$$M(\lambda) = i(\lambda)/(S(\lambda)O(\lambda)D(\lambda)).$$

8.5 Measurement Procedures

From figure 8.6, we see that we can determine our input light source $S(\lambda)$ from our measured detector current $i(\lambda)$ through the use of the equation

$$i(\lambda) = S(\lambda)O(\lambda)M(\lambda)D(\lambda).$$

However, it would be almost impossible to obtain accurate values using this equation since the terms $O(\lambda)$, $M(\lambda)$, and $D(\lambda)$ are very difficult to obtain.

The solution to this problem is to use a light source for which the output is already known. This is a calibrated or standard source. We may then perform the following sequence of measurements:

1. Measure the standard source:
$$i_S(\lambda) = S_S(\lambda)O(\lambda)M(\lambda)D(\lambda)$$
2. Measure the test source:
$$i_T(\lambda) = S_T(\lambda)O(\lambda)M(\lambda)D(\lambda)$$
3. Take the ratio of the two measurements. Note that if you have been very careful not to change anything but the two sources, the difficult terms $O(\lambda)$, $M(\lambda)$, and $D(\lambda)$, cancel. We are thus left with
$$i_S(\lambda)/i_T(\lambda) = S_S(\lambda)/S_T(\lambda).$$

From this $S_T(\lambda)$ may be calculated, since the other three terms are known.

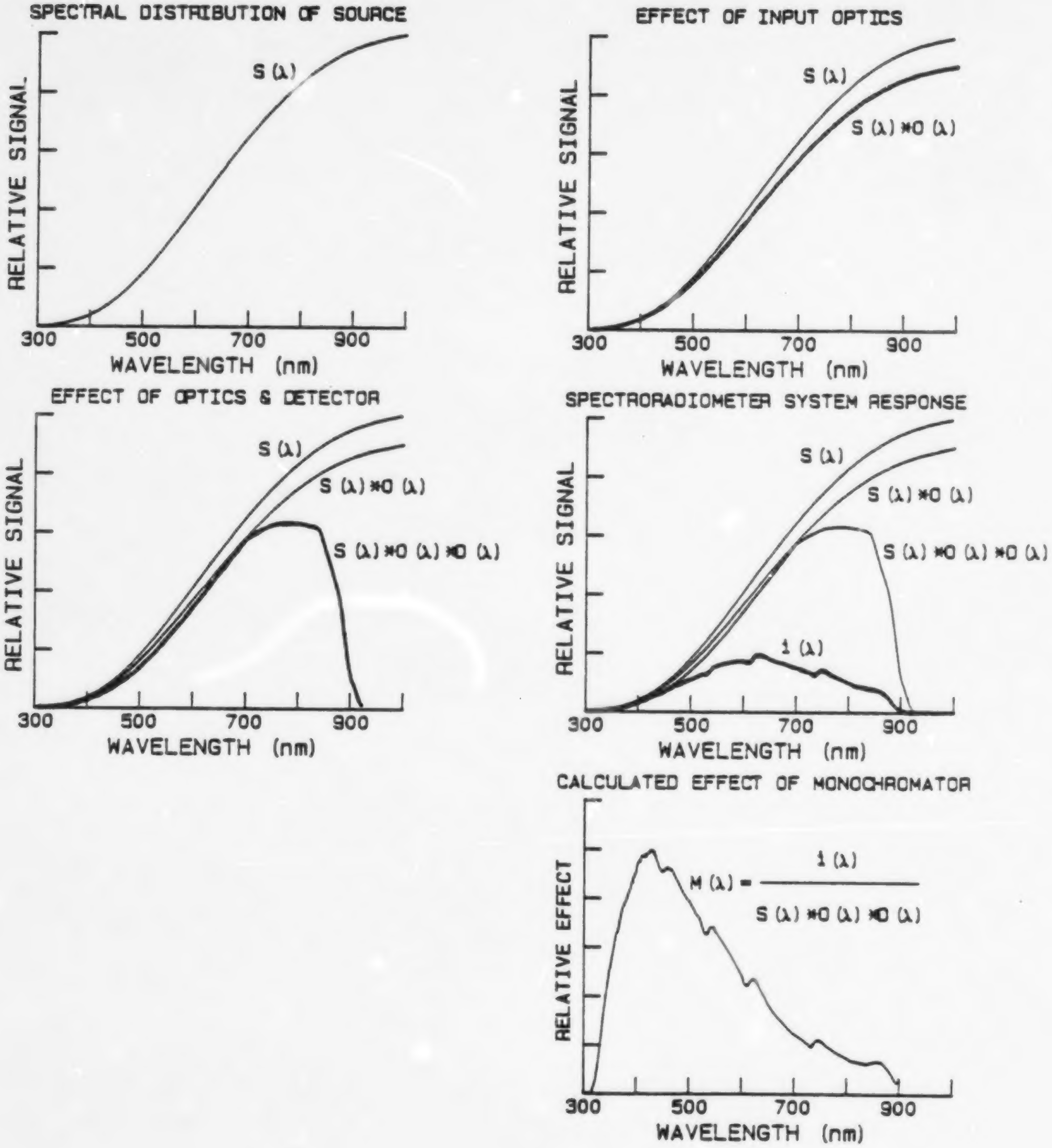

Figure 8.7. Effects of the spectroradiometer components on the light emitted by a source.

8.6 Automated Spectroradiometers

An automated system may be constructed as shown in figure 8.8.

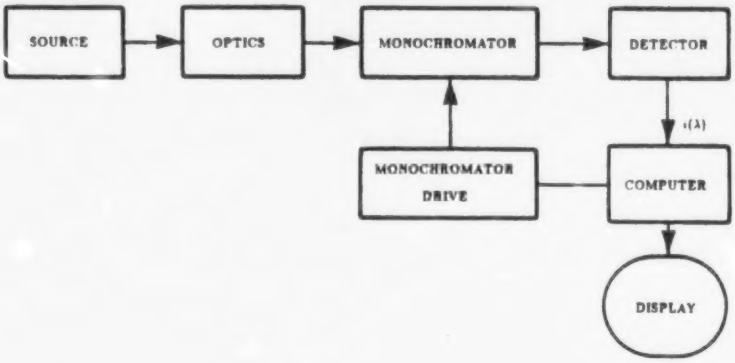
SOURCE
OPTICS
MONOCHROMATOR
DETECTOR
i(λ)
MONOCHROMATOR DRIVE
COMPUTER
DISPLAY

Figure 8.8. Components of an automated spectroradiometer system.

In addition to the ease with which this type of system may be used, there is usually the 'advantage' that the spectrum of the desired source may be obtained without measurement of a standard source. This is because the manufacturer has performed that measurement for you and stored the data in the computer. There are two very critical assumptions which are therefore being made when you use such a system:

1. The correct calibration configuration was used.
2. The system has a long term stability.

8.7 Applications

8.7.1 Spectral irradiance measurement

The use of a spectroradiometer for a spectral irradiance measurement is illustrated in figure 8.9. The diffuser, lens, and spectroradiometer are all considered as part of the measurement device. Therefore, they should all remain in exactly the same position for the measurement of the test and the standard lamp.

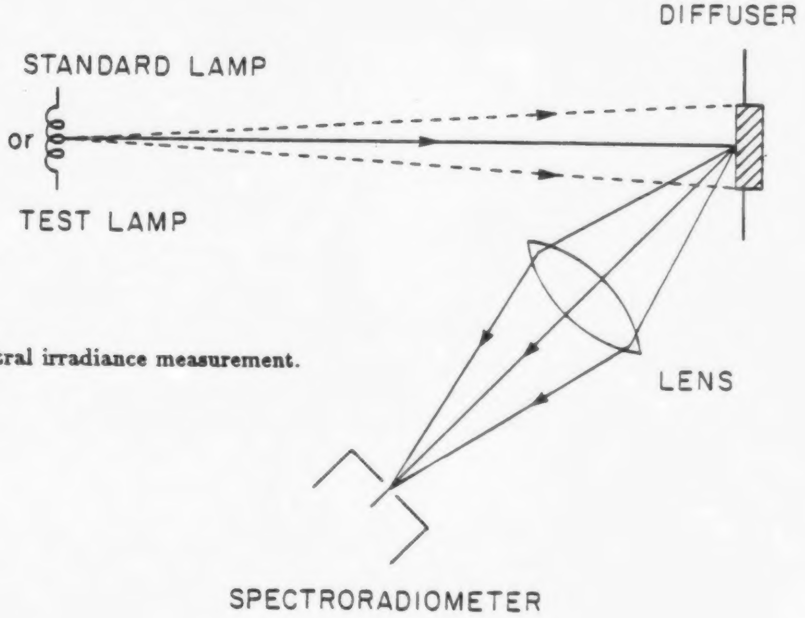
DIFFUSER
STANDARD LAMP
or
TEST LAMP
LENS
SPECTRORADIOMETER

Figure 8.9. A spectral irradiance measurement.

8.7.2 Spectral radiance measurement (ribbon filament)

The measurement of spectral radiance requires, in addition to the spectral considerations already discussed, rather rigid constraints in the input optics as well. This is due to the definition of radiance (see lecture 7) and is illustrated in figure 8.10.

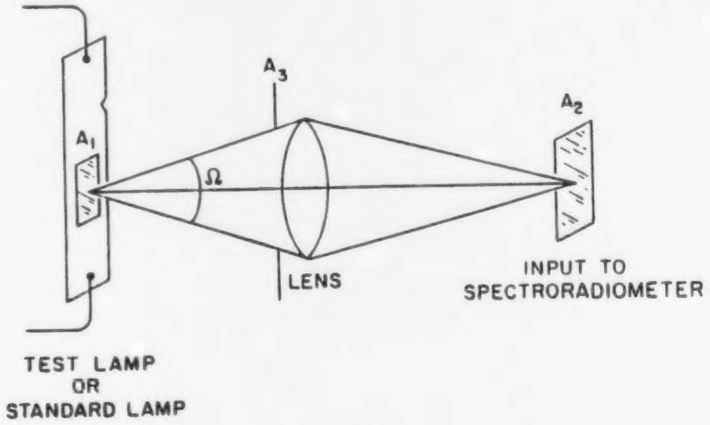

Figure 8.10. Input optics required for a spectral radiance measurement. A_1 is the defined area on the standard lamp. A_2 is the image of A_1. The input to the spectroradiometer must be modified to accept only light from this image. Ω is the defined solid angle for the standard lamp. A_3 is an aperture constructed to pass only light from the defined solid angle Ω.

8.8 References

Grum, F., and Becherer, Richard J., 1979, *Optical Radiation Measurements, Volume 1, Radiometry*, Academic Press: New York.

NBS Technical Notes 910-1 to 910-11, *Self-study Manual on Optical Radiation Measurement*, U.S. Department of Commerce: Washington, D.C.

Sawyer, R.A., 1963, *Experimental Spectroscopy*, 3rd Edition, Dover Publications Inc., New York.

Notes

Lecture 9: J. C. Zwinkels

SPECTROPHOTOMETRY

9.1 Introduction

Spectrophotometry involves a comparison at each wavelength of the radiant power leaving an object to that incident on it. The quantities measured are the spectral transmittance for transparent objects and the spectral reflectance factor for opaque objects. Since the measurement is relative, it is independent of the spectral power distribution of the source.

The spectral transmittance, $\tau(\lambda)$ is the ratio of the transmitted radiant flux to the incident radiant flux over a narrow wavelength band (instrument spectral bandpass).

The spectral reflectance factor, $\beta(\lambda)$ is the ratio of the flux reflected from a specimen to the flux reflected from a standard surface, the so-called perfectly reflecting diffuser, under the same geometric conditions of measurement. If the viewing condition is unidirectional, it is referred to as the spectral radiance factor, whereas if the viewing condition is hemispherical (over 2π), it is called the spectral reflectance and denoted by $\rho(\lambda)$.

9.2 Instrumentation

The key components of a spectrophotometer are shown in figure 9.1. Most commercial instruments are of this double-beam ratio-recording type, where both the sample and reference signals are alternately monitored at each wavelength. Its advantages include compensation for fluctuations in source output, detector response, or atmospheric absorption, as well as, the capability of extending the dynamic range of the photometric scale by reference beam attenuation.

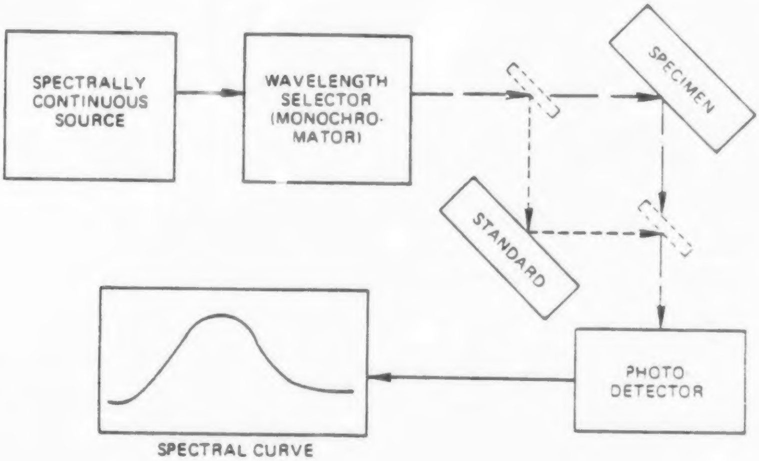
SPECTRALLY CONTINUOUS SOURCE
WAVELENGTH SELECTOR (MONOCHRO-MATOR)
SPECIMEN
STANDARD
PHOTO DETECTOR
SPECTRAL CURVE

Figure 9.1. Block diagram of the basic components of a spectrophotometer.

The lamp sources and photodetectors commonly employed in spectrophotometry are listed in table 9.1.

Sources	Range of Use
Deuterium lamp	180–400 nm
Tungsten–halogen lamp	320–3200 nm
Xenon arc lamps	200–800 nm
Detectors	
Photomultipliers	180–850 nm
Silicon photodiodes	250–1100 nm
Lead sulphide	800–3200 nm

Table 9.1. Sources and Detectors Used in Spectrophotometry.

Wavelength separation is achieved with a monochromator. An abridged spectrophotometer uses a series of narrow passband interference filters or wedges to isolate different bands of wavelength. A single monochromator, as shown in figure 9.2, employs a single dispersing element, either a prism or grating, to give a continuous wavelength separation. For a very low stray light requirement (<0.001%T), a double monochromator is used.

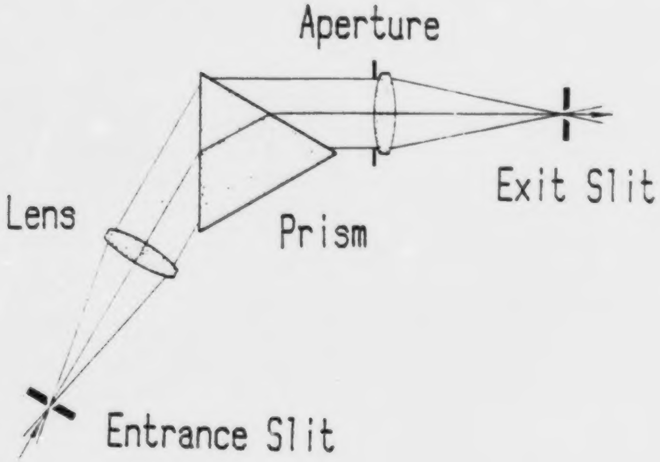

Figure 9.2. Schematic diagram of a single-prism monochromator.

The two most common monochromator optical configurations are shown in figure 9.3. The Czerny-Turner system (a) uses two mirrors whereas the Ebert (b) uses a single spherical-concave mirror to first collimate and then refocus the light.

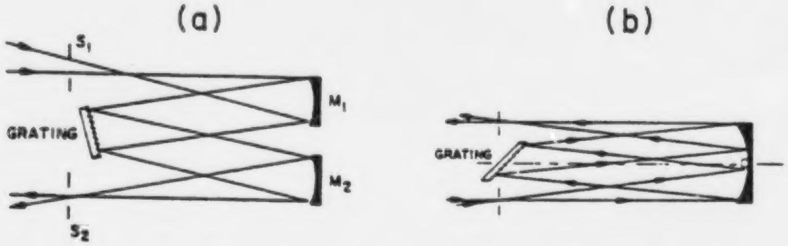

Figure 9.3. Optical configurations: (a) Czerny-Turner, (b) Ebert.

9.3 Monochromator Parameters

The operation of a monochromator requires consideration of three key performance parameters: resolution, throughput, and stray light. Since they are interdependent, trade-offs must be made for best performance for a particular application.

Reciprocal Linear Dispersion, D^{-1}

How well the monochromator spreads the spectrum over the focal plane of the exit slit.

$$D^{-1}(\mathrm{nm/mm}) \propto 1/n; n = \text{groove density (g/mm)}$$

Resolution, $\Delta\lambda$

the minimum detectable difference between peaks.

$$\Delta\lambda(\mathrm{nm}) = w(\mathrm{mm}) \times D^{-1}(\mathrm{nm/mm}); w = \text{slit width}$$

Throughput, TP

overall effectiveness with which a monochromator transmits radiation.

$$TP \propto Dh/(f/no)^2; h = \text{slit height}$$

Monochromator Stray Light, MSL

undesired radiation wavelengths.

$$MSL \propto h: \quad \text{(single-monochromator)}$$

$$MSL \propto wh^2: \quad \text{(double-monochromator)}$$

9.4 Reflection and Transmission Phenomena

When light impinges upon a material, several phenomena can occur as shown in figure 9.4. In order to completely specify the appearance of an object (color and geometric attributes), it is necessary to identify both the measurement scale (transmittance or reflectance, preceded by an adjective, such as specular, diffuse, regular, etc.), as well as the geometric conditions (angle of incidence and viewing, incidence and receptor field angles).

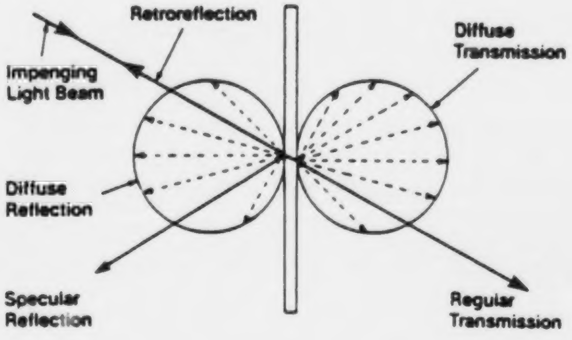
Retroreflection
Impinging Light Beam
Diffuse Reflection
Specular Reflection
Diffuse Transmission
Regular Transmission

Figure 9.4. Idealisations of reflection and transmission phenomena, showing components.

Some of the CIE recommended measurement geometries are illustrated in figure 9.5. The diffuse illumination or viewing conditions are achieved with an integrating sphere coated with an efficient white diffuser, such as $BaSO_4$.

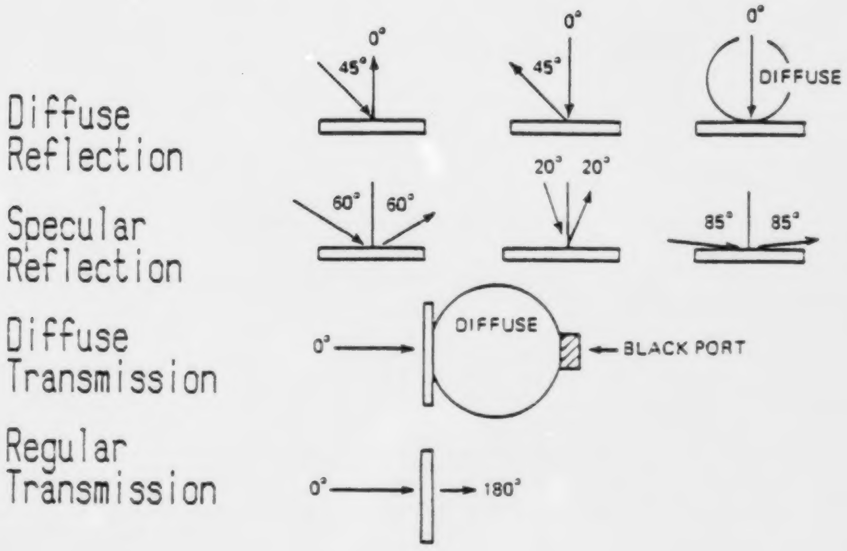
Diffuse Reflection
0°
45°
0°
45°
0°
DIFFUSE
Specular Reflection
60°
60°
20°
20°
85°
85°
Diffuse Transmission
0°
DIFFUSE
BLACK PORT
Regular Transmission
0°
180°

Figure 9.5. Axial directions of illumination and view commonly used in spectrophotometry.

9.4.1 Regular Transmission

The regular transmittance, T, is the ratio of the flux transmitted through the specimen to the flux measured through clear air.

The internal transmittance, τ, is the ratio of the flux reaching the exit surface of the specimen to the flux that penetrates the entry surface, and for a homogeneous, isotropic absorbing material, is described by Beer's law:

$$\tau_\lambda = 10^{-\varepsilon_\lambda cd}$$

where ε_λ is the absorption coefficient at wavelength λ, c is the concentration, and d is the thickness. It is related to the regular transmittance, T, by the expression:

$$T = (1 - R)^2 \tau$$

where R is the Fresnel reflectance.

The optical density, D, is the negative logarithm to the base ten of the measured transmittance, $D = -\log_{10}(T)$, whereas the absorbance, A, is the negative logarithm to the base ten of the internal transmittance, $A = -\log_{10}(\tau) = \varepsilon cd$.

In the case where the sample has plane-parallel surfaces so that multiple internal reflections can occur, the regular transmittance T is given by:

$$T = \frac{(1-R)^2 \tau}{1 - (R\tau)^2}.$$

Tilting the sample minimizes interreflections but introduces obliquity errors (Mielenz and Mavrodineau, 1973).

9.4.2 Specular Reflection

Specular reflection is mirror-like reflection for which the angle of reflection equals the angle of incidence. Relative reflectance measurements are made against calibrated mirrors, such as Ag, Al (NBS SRM's 2003a, 2023, 2024), Au (NBS SRM 2011), and Rh. Absolute reflectance measurements can be made with a 'VW' optical accessory illustrated in figure 9.6. The ratio of the radiant flux P_2 to the radiant flux P_1 gives the square of the sample reflectance directly.

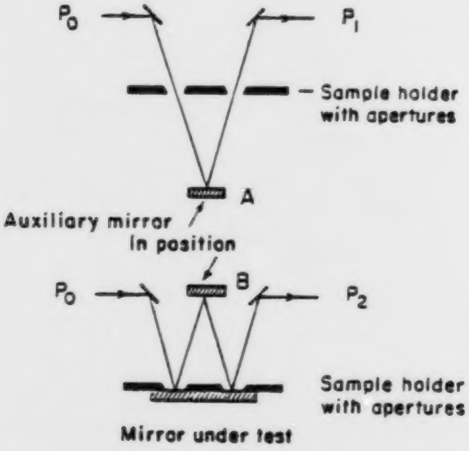

Figure 9.6. Specular reflectance accessory for absolute measurements using the 'VW' configuration of Strong.

9.4.3 Diffuse Transmission

Diffuse transmittance is measured with normal incidence and diffuse viewing, using the same sphere accessory as for 0°/d reflectance measurements. The sample is placed normally to the entrance port of the sphere and a white plate of the same material as the sphere coating is placed in the usual sample port.

Linearity errors at the end-points of the photometric scale (0% and 100% transmittance) are checked for by first recording the spectrum with an opaque sample, $I_0(\lambda)$ and then with no sample in the beam, $I_{100}(\lambda)$. The sample signal, $I_s(\lambda)$ is then corrected, according to the following expression, to give the sample diffuse transmittance, $\tau_s(\lambda)$:

$$\tau_s(\lambda) = \frac{I_s(\lambda) - I_0(\lambda)}{I_{100}(\lambda) - I_0(\lambda)}.$$

9.4.4 Diffuse Reflection

Diffuse reflection is reflection in which the flux is scattered in directions other than specular, whereas total reflection includes the flux scattered in all directions. To convert an integrating sphere from total to diffuse-only geometry, the part of the sphere wall which is normally hit by the specularly reflected light, is replaced by a black plug or light trap (see figure 9.7).

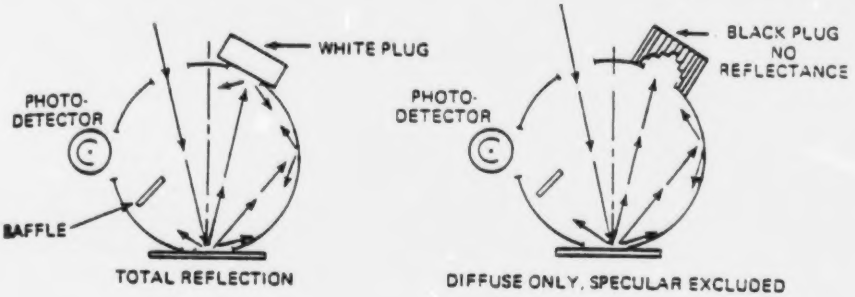

Figure 9.7. Converting a sphere from geometry for measurement of total reflection to that for diffuse only, with specular excluded.

The diffuse reflectance of a sample is obtained by a procedure similar to that described for diffuse transmittance, where $I_0(\lambda)$ is recorded with a black cavity, and $I_{100}(\lambda)$ is recorded with a white reflectance standard; $\rho_{100}(\lambda)$ is the absolute reflectance of the white standard.

$$\rho_s(\lambda) = \left(\frac{I_s(\lambda) - I_0(\lambda)}{I_{100}(\lambda) - I_0(\lambda)}\right) \rho_{100}(\lambda)$$

9.5 White Reflectance Standards

In 1971, the CIE adopted the perfectly reflecting diffuser as the primary standard for diffuse reflectance measurements. It is defined as the ideal uniform diffuser according to Lambert's law with a reflection equal to unity. However, no material standard satisfies these requirements so, in practice, we use transfer standards or working standards.

A transfer standard is used to transfer a calibration from one instrument to another and should be very matte, highly opaque, spectrally non-selective, and have a high reflectance value; examples are: MgO, $BaSO_4$, Halon, matte opal glass.

A working standard is used for day-to-day calibrations on the same instrument and should be very stable and cleanable; examples are: glossy opal glass, Vitrolite, and white enamel plates.

Figure 9.8 shows absolute spectral reflectance curves of some of these white standards.

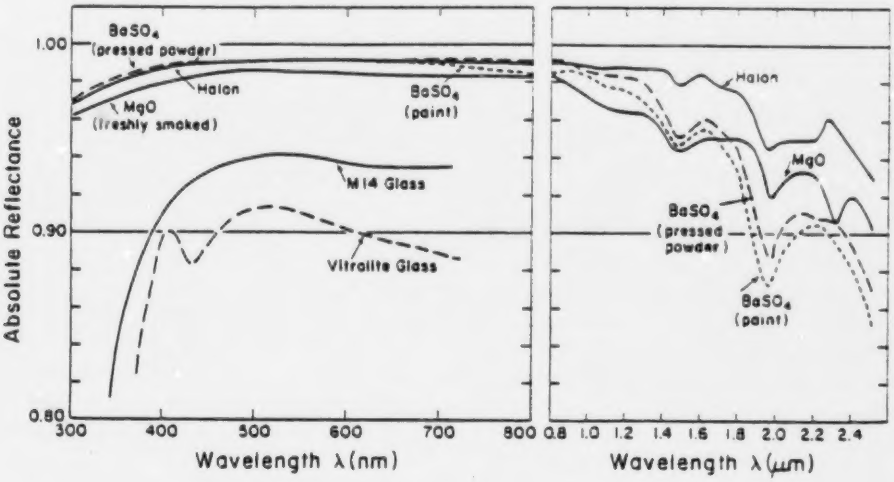

Figure 9.8. Absolute spectral reflectance curves of different materials used as transfer or working standards of reflectance.

9.6 Diagnostic Error Checks

9.6.1 Instrument Stability

Instrument stability should be periodically checked with a set of neutral glass filters, such as NBS SRM 930. A control chart will immediately indicate any malfunction of the spectrophotometer.

9.6.2 Wavelength Scale

The most accurate method for checking the wavelength scale is to use line source standards. A low-pressure Hg lamp in a quartz envelope is the best single source for wavelength calibration in the UV-VIS range. In a spectrophotometer equipped with a deuterium source, the deuterium emission lines at 486.0 and 656.1nm are also useful for a quick check.

Material standards such as didymium and holmium oxide glasses, which exhibit many sharp transmission features in the visible spectrum, are more convenient for use in commercial instruments. However, the positions of their transmission minima are very sensitive to the width of the spectral passband of the spectrophotometer, so care must be exercised to use the same spectral bandwidth with which the filter was originally calibrated.

9.6.3 Linearity

Photometric accuracy requires linearity of the photometric scale, that is, that the output quantity of the detector is exactly proportional to the input quantity over a specified range. In a standards lab, linearity checks are based on the light addition principle using multiple sources (Reule, 1968) or multiple apertures (Mielenz and Eckerle, 1972).

In practice, secondary standards are used in the form of calibrated glass filters (NBS SRM 930), metal-on-quartz filters (NBS SRM 2031), or precisely prepared solutions of carefully measured chemical reference materials (NBS SRM 931b). Linearity tests based on Beer's law (sample absorbance linearly related to concentration) are not sensitive to exponential non-linearity, ie. the possibility that the measured transmittance values all have exactly the same proportional error (Hawes, 1971), so they must be used in combination with a second independent linearity check.

9.6.4 Stray light

The stray light error is defined as the ratio of the spurious radiation (as detected) to the primary or desired radiation (as detected), and primarily arises from inter-reflections and scattering inside the monochromator. Since stray light is defined as a ratio, its effects are greatest at the end of a monochromator's source energy and detector response limit, where the primary radiation is low. This is why instrument manufacturers usually specify the stray light at 220nm (where the D_2 arc energy is low) and at 340nm (close to the minimum of the tungsten lamp energy).

Methods for Estimating Stray Light Error:

a) Sharp Cutoff Filters

This method involves measuring the absorbance of a filter just beyond its cut-off point. The filter must have an absorbance curve which rises sharply over a narrow wavelength range. Figure 9.9 illustrates the principle of the method. The solid curve represents the true absorbance of the filter and the dashed curve is the apparent absorbance in the presence of stray light. The stray light error at wavelength λ' is given by $x = 10^{-A'}$. Table 9.2 lists several filters recommended by the American Society for Testing and Materials (ASTM) for checking stray light performance.

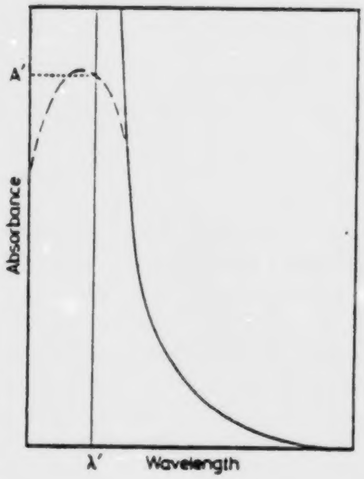

Figure 9.9. The use of a cutoff filter to measure stray light.

Material	Concentration	Path Length	Spectral Range*
KCl (aq)	12g/lt	1.00 cm	175-200 nm
NaBr (aq)	10g/lt	1.00 cm	195-223 nm
NaI (aq)	10g/lt	1.00 cm	210-259 nm
acetone	100%	1.00 cm	250-320 nm
$NaNO_2$ (aq)	50g/lt	1.00 cm	300-385 nm

Table 9.2. Cutoff Filters for Tests of Stray Radiant Energy.

*The longer wavelength gives 10^{-4} transmittance point.

b) Deviations from Beer-Lambert Law

These methods involve either measuring the absorbance of a series of solutions of different concentrations, keeping the path length constant, or varying the path length and using the same solution for all measurements. Figure 9.10 illustrates how the presence of stray light can give low apparent absorbance values (negative deviations from Beer's law) and how the effect increases with increasing sample absorbance.

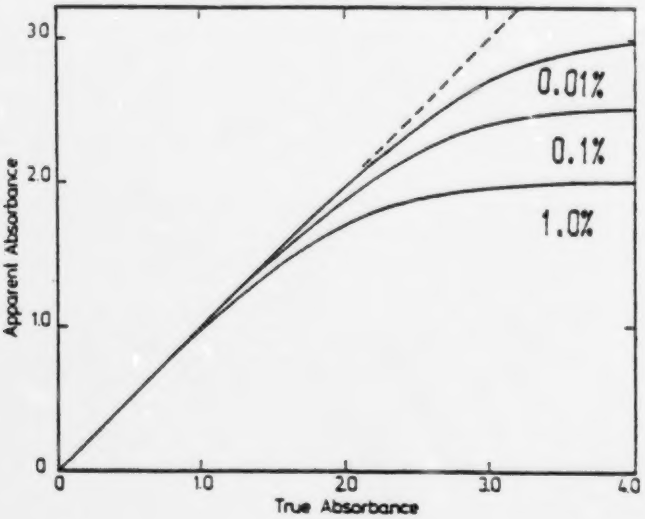

Figure 9.10. Plots of apparent absorbance against true absorbance for different percentages of stray light.

9.6.5 Fluorescence

Sample fluorescence occurs when the sample absorbs radiant flux at one wavelength and reemits it at another, usually longer wavelength. Its presence can be detected by viewing the sample under a UV lamp and observing a violet-blue color.

Figure 9.11 shows both incorrect (a) and correct (b,c) procedures for measuring fluorescent materials. In the usual monochromatic mode (a), light emitted by sample fluorescence will erroneously be attributed to the shorter wavelengths exciting the fluorescence.

In the polychromatic mode (b), where the monochromator is placed between the sample and the photodetector, the fluorescence is correctly attributed to the longer wavelengths where it is emitted. However, the shape of the spectral curve is very dependent on the spectral power distribution of the source so in order to correlate spectrophotometric measurements of fluorescent materials with visual perception, it is necessary that the instrument source resemble the illuminant to be used for color evaluation (eg. a high-pressure Xe arc lamp is an excellent 'daylight simulator').

In the two-monochromator mode (c), the sample is exposed to monochromatic radiant flux of different wavelengths and the corresponding emission band is recorded. In principle, it is the most accurate method for characterizing fluorescent materials, but it is very involved and time-consuming.

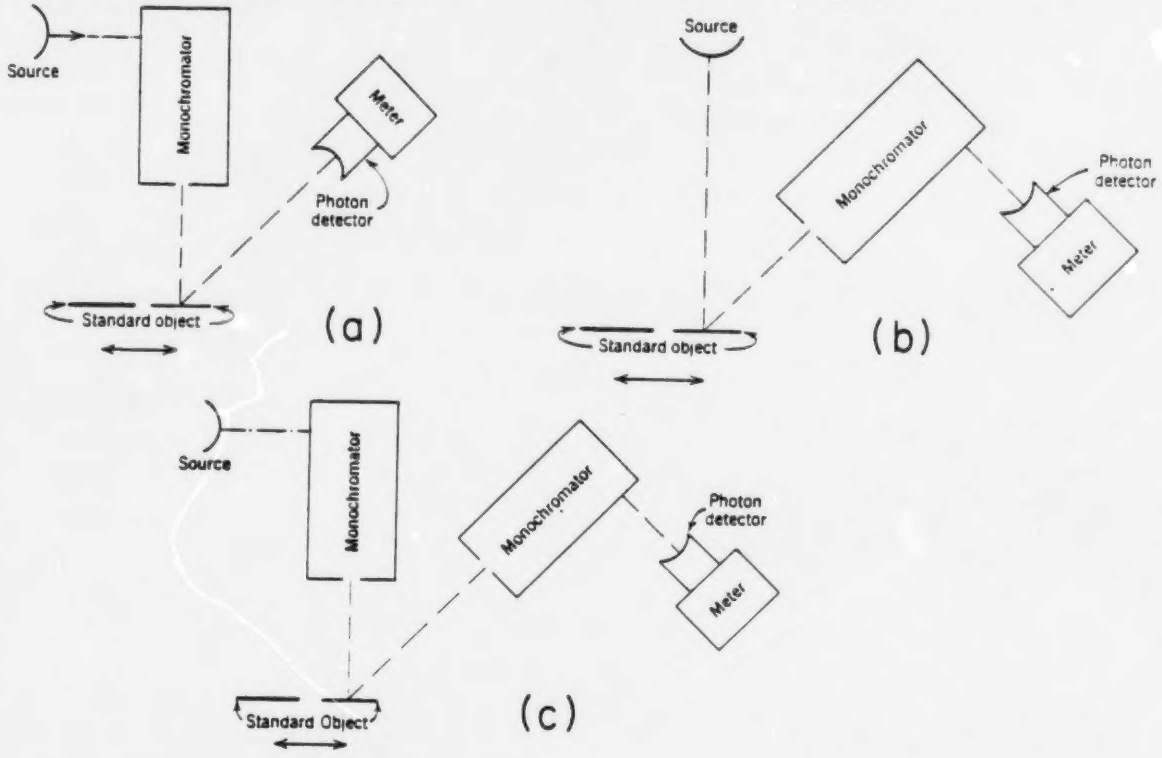
Source
Monochromator
Meter
Photon detector
Standard object
(a)
Source
Monochromator
Photon detector
Meter
Standard object
(b)
Source
Monochromator
Monochromator
Photon detector
Meter
Standard Object
(c)

Figure 9.11. Schematic diagram illustrating (a) incorrect and (b,c) correct experimental arrangements for measuring fluorescent objects.

9.6.6 Translucence

Translucence is a property of a sample by which a major portion of the incident light undergoes internal scattering causing losses out the back and edges of the sample. Its presence is indicated by the measured spectral reflectance or transmittance factors changing with varying beam diameter or sample thickness.

The diffuse reflectance of a translucent sample, such as Vitrolite, should be measured with a black (1% R) backing material, and with the beam diameter 50% of the port width or height (to avoid edge losses).

9.7 Operating Parameters

Poor selection of instrument operating parameters, such as slit width, scan speed, response time, and slit height can contribute to spectrophotometric errors, such as wavelength shifts, peak distortion, poor resolution, and transmittance errors. Table 9.3 summarises the effects of changing these parameters to either side of the optimum value.

Parameter	Selection	Effect
Slit width	too narrow	increased noise
	too wide	loss of resolution, transmittance errors
Scan speed	too fast	loss of resolution, wavelength errors
	too slow	increased noise, measurement time
Response time	too short	increased noise
	too long	reduced sensitivity
Slit height	too large	increased stray light
	too small	increased noise

Table 9.3. Effect of Operating Parameters on Spectrophotometric Performance.

Another instrumental factor affecting SNR in double-beam spectrophotometers is attenuation of the reference beam. At low sample absorbances this technique is not useful, but at high sample absorbances, it gives a significant improvement in SNR, as can be seen from figure 9.12.

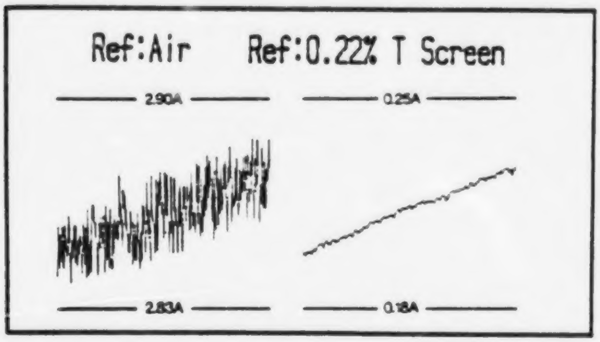

Figure 9.12. Measured SNR of a strongly absorbing sample both without, and with reference beam attenuation.

9.8 References

White Reflectance Standards

Erb, W. and Budde, W., 1979, 'Properties of Standard Materials for Reflection', *Color Res. Appl.*, **4:** 113-118.

Grum, F. and Wightman, T.E., 1977, 'Absolute Reflectance of Eastman White Reflectance Standard', *Appl. Opt.*, **16:** 2775-2776.

Weidner, V.R., 1983, 'White opal Glass Diffuse Spectral Reflectance Standards for the Visible Spectrum (SRM's 2015 and 2016)', *NBS Special Publication*, 260-82.

Weidner, V.R. and Hsia, J.J., 1981, 'Reflection Properties of Pressed Polytetrafluoroethylene Powder', *J. Opt. Soc. Am.*, **71:** 856-861.

Diagnostic Error Checks

Burgess, G. and Knowles, A., 1981, *Standards in Absorption Spectrometry*, London, Chapman and Hall, Ltd.

Habell, K.J. and Jackson, A., 1963, 'Spectral Lines for Refractometry', *Optica Acta*, **10:** 217-221.

Hawes, R.C., 1971, 'Technique for Measuring Photometric Accuracy', Appl. Opt., *10:* 1246-1253.

McNeirny, J. and Slavin, W., 1962, 'A Wavelength Standard for

Ultraviolet-Visible-Near Infrared Spectrophotometry', *Appl. Opt.*, **1:** 365-367.

Mielenz, K.D. and Eckerle, K.L., 1972, 'Spectrophotometer Linearity Testing Using the Double-Aperture Method', *Appl. Opt.*, **11:** 2294-2303.

Mielenz, K.D. and Mavrodineau, R., 1973, 'Reflection Correction for High Accuracy Transmittance Measurements on Filter Glasses', *J. Res. Nat. Bur. Stand. (U.S.)*, **77A:** 86-90.

Poulson, R.E., 1964, 'Test Methods in Spectrophotometry: Stray Light Determination', *Appl. Opt.* **3:** 99-104.

Reule, A., 1968, 'Testing Spectrophotometer Linearity', *Appl. Opt.*, **7:** 1023-1028.

Reule, A.G., 1975, 'Errors in Spectrophotometry and Calibration Procedures to Avoid Them', *NBS Special Publication 466*, 99-114.

Sharpe, M.R. 1984, 'Stray Light in UV-VIS Spectrophotometers', *Anal. Chem.*, **56:** 339A-356A.

Venable, W.H. and Eckerle, K.L., 1979, 'Didymium Glass Filters for Calibrating the Wavelength Scale of Spectrophotometers - SRM 2009, 2010, 2013 and 2014', *NBS Special Publication*, 260-66.

Weidner, V.R., Mavrodineau, R., Mielenz, K.D., Velapoldi, R.A., Eckerle, K.L. and Adams, B., 1985, 'Spectral Transmittance Characteristics of Holmium Oxide in Perchloric Acid Solution', *J. Res. Nat. Bur. Stand. (U.S.)*, **90:** 115-125.

Notes

Notes

Lecture 10: A. R. Robertson

REFLECTANCE STANDARDS

10.1 Primary Standard

The primary standard for measurements of reflectance factor is the so-called 'perfect reflecting diffuser'. The perfect reflecting diffuser has a reflectance factor of unity at all wavelengths and it diffuses incident light perfectly so that the reflected radiance is the same in all directions. Such a diffuser does not exist in practice so it is a theoretical concept only.

The reflectance factor of real samples can be measured relative to the perfect reflecting diffuser by comparing the incident light with the reflected light. However this is a very difficult measurement to make accurately so in practice secondary standards are calibrated by a few standardizing laboratories and distributed to other laboratories. These laboratories can then measure their samples by comparatively simple relative measurements in which the light reflected from the sample is compared with the light reflected from the calibrated secondary standard under identical conditions.

10.2 NRC Absolute Reflectometer

The reflectometer used at NRC for the calibration of secondary standards against the perfect diffuser is shown in figure 10.1.

Light from a tungsten-filament lamp passes through a filter and is focussed on the wall of a 20 cm diameter integrating sphere. The filter is one of a set of interference filters in a filter wheel to allow spectral measurements. A baffle near the centre of the sphere blocks any direct radiation to the sample from the image on the sphere wall. A photometer system, consisting of a lens, opal glass and photomultiplier measures the radiance L_s of the sample and then swings round to measure the radiance L_w of the sphere wall. It can be shown that the reflectance, ρ , of the sample is given by

$$\rho = (L_s/L_w) \cdot (A_0/(A_1 + A_2))$$

where A_0 is the total sphere area ($= 4\pi r^2$), A_1 is the area of the sample port, and A_2 is area of the remaining sphere wall.

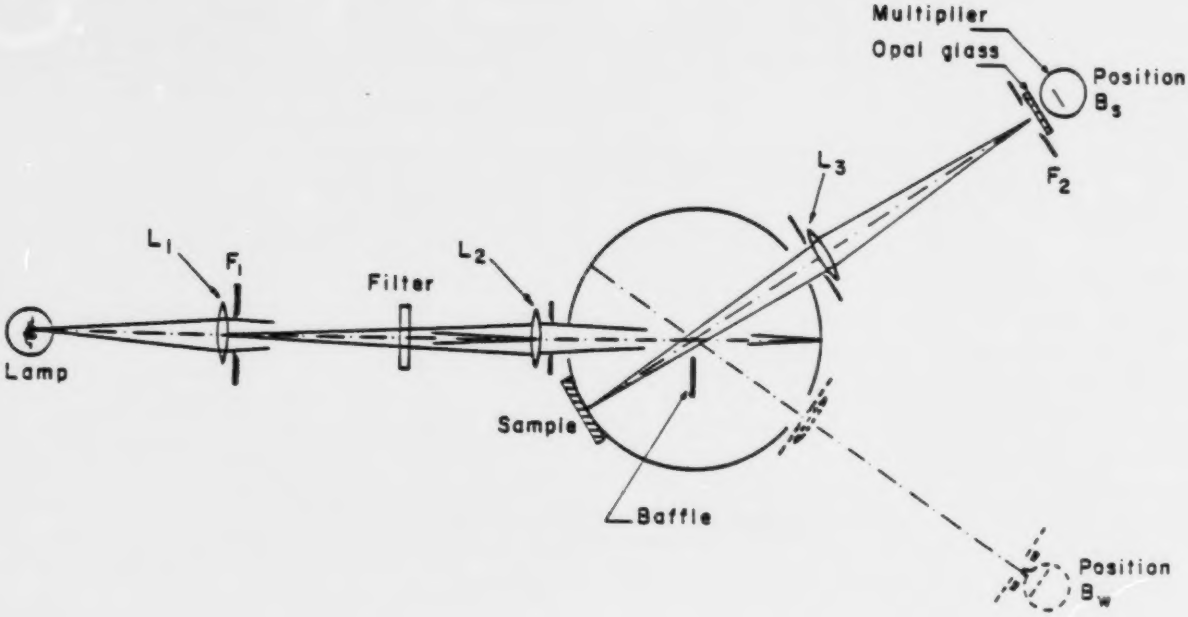

Figure 10.1. NRC Absolute Reflectometer (Budde, 1976).

10.3 Reflectance Factors of Pressed $BaSO_4$

Tablets made by pressing barium sulphate ($BaSO_4$) powder in a specially designed press are among the whitest materials available. Spectral reflectance factors of two types of $BaSO_4$ are shown in table 10.1. These are very pure grades of $BaSO_4$ made specifically for reflectance factor standards.

Wavelength (nm)	Merck $BaSO_4$	Eastman $BaSO_4$
358	0.949	0.979
388	0.971	0.984
453	0.984	0.989
509	0.986	0.991
559	0.985	0.991
639	0.987	0.992
699	0.987	0.991
751	0.982	0.992

Table 10.1. Absolute spectral reflectance factors (d/0) of pressed $BaSO_4$ powders (Budde and Dodd, 1970).

The measurements shown in table 10.1 were for d/0 geometry. The values are slightly different for other geometries. For example some values for 45/0 geometry measured on two samples of Merck $BaSO_4$ are shown in table 10.2.

Wavelength (nm)	Sample 1	Sample 2
394	0.999	0.993
403	1.002	1.000
435	1.007	1.006
467	1.012	1.009
495	1.015	1.011
527	1.014	1.010
545	1.014	1.010
578	1.015	1.010
606	1.017	1.012
661	1.015	1.010
708	1.016	1.010
772	1.018	1.012

Table 10.2. Absolute spectral reflectance factors, (45/0) of two samples of Merck $BaSO_4$ (Erb, 1971).

10.4 Transfer and Working Standards

It is important to distinguish between two types of reflectance factor standards.

Transfer standards are used to transfer a calibration from one instrument (such as an absolute reflectometer) to another (such as a spectrophotometer). It is important that they are not sensitive to small geometrical differences between the instruments so they must be good diffusers, opaque, and non-selective in wavelength. Since they need only be used once in each instrument it is not necessary that they be rugged and cleanable.

Working standards, on the other hand, are for day-to-day use in a single instrument. They must be rugged and cleanable but good diffusion and opacity are less important.

A list of the requirements for both types of standards is given in table 10.3. Suitable materials for both types are listed in table 10.4. The spectral reflectance factors of some of the materials are shown in figure 10.2.

Property	Transfer Standard	Working Standard
Good mat diffuser	+	−
Glossy surface	−	+
High reflectance factor	+	0
Non-selective	+	0
Opaque	+	0
Uniform	+	+
Flat	+	+
Stable (days or weeks)	+	+
Stable (months or years)	0	+
Transportable	+	0
Cleanable	0	+
Non-fluorescent	+	+
Non-hygroscopic	+	+
Non-thermochromic	+	+

Table 10.3. Necessary and desirable properties of reflectance factor standards. (+: necessary; 0: desirable, but not necessary; −: undesirable)

Material	Application
Pressed MgO tablet	T
Pressed $BaSO_4$ tablet	T
Painted $BaSO_4$ panel	W
Opal glass MS 20, mat surface	T , W
Opal glass MS 20, glossy surface	W
Ceramic tile (mat surface)	T
Vitrolite glass (polished or mat)	W
Enamel panels (glossy)	W
Polytetrafluorethylene ('Halon')	T

Table 10.4. Suitable materials for reflectance factor standards.
T: transfer standard; W: working standard

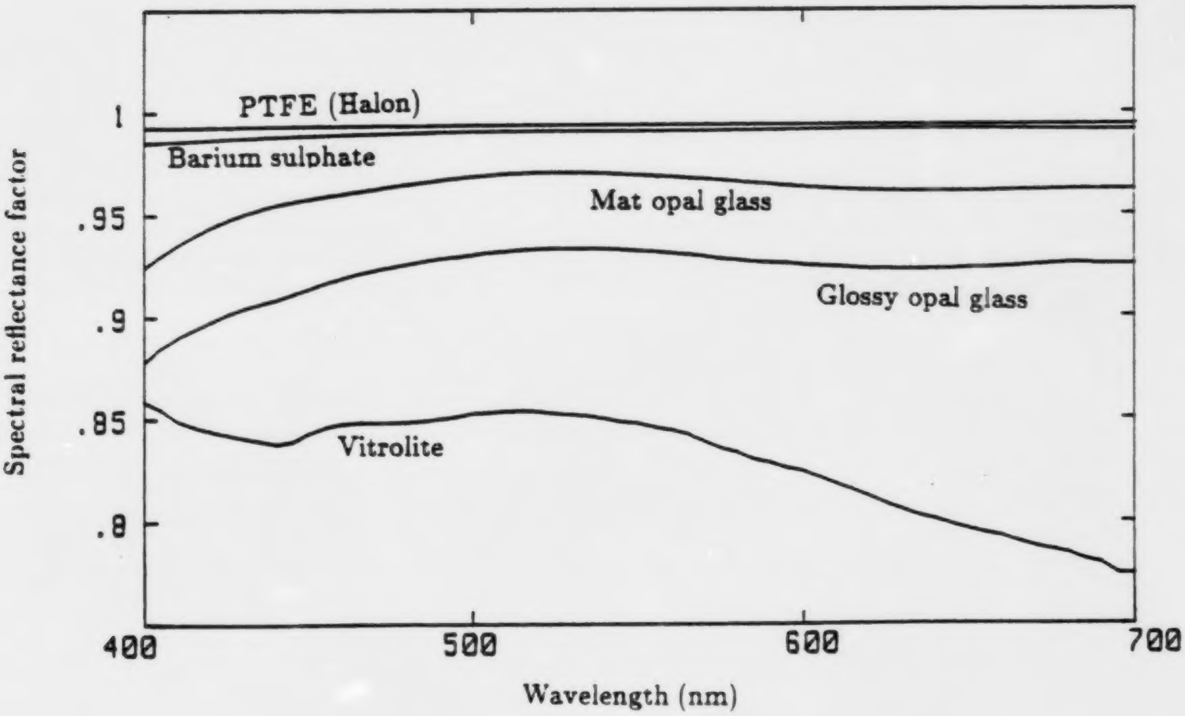

Figure 10.2. Spectral reflectance factors of some commonly used standards.

10.5 References

Budde, W., 1976, 'Calibration of Reflectance Standards', *J. Res. Natl. Bur. Stds.*, **80A**: 585-595.

Budde, W. and Dodd, C.X., 1970, 'Absolute Reflectance Measurements in the d/0 Geometry', *Die Farbe*, **19**: 94-102.

Erb, W., 1971, 'Bestimmung des spektralen Leuchtdichtekoeffizienten für die Messgeometrie 45/normal', *Die Farbe*, **20**: 179-186.

Notes

Notes

Lecture 11: A. R. Robertson

TRISTIMULUS COLORIMETRY

11.1 The Principle of Tristimulus Colorimetry

A tristimulus colorimeter is an instrument with spectral response functions directly proportional to the colour-matching functions of one or the other CIE standard colorimetric observers. This feature enables direct measurement of CIE tristimulus values without computation.

The fundamental problem in building a tristimulus colorimeter is to adjust three photodetectors so that, to a sufficient approximation, their responses are proportional throughout the visible spectrum to the desired colour-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ respectively, or to some linear combination of them. Good references for the subject are Billmeyer and Saltzman (1981), Judd and Wyszecki (1975) and Wyszecki and Stiles (1982).

11.2 Matching Techniques

11.2.1 Full Filtering

One method to modify the spectral responsivity of a photodetector is to select coloured filters of suitable spectral transmittance and to combine them in series in front of the detector. In this arrangement, light is filtered successively by each filter before it falls on the detector. Figure 11.1 illustrates what can be achieved by this method. Crosses show the photodetector responsivity, the full line is the $V(\lambda)$ curve, and the circles indicate the fit achieved by placing several selected filters in series in front of the detector. This is a particularly good fit, achieved with great care at NRC. Many commercial $V(\lambda)$ fits are not as good as this.

To fit the colour-matching functions well requires four such filter combinations because the double-peaked $\bar{x}(\lambda)$ curve has to be fitted in two parts. Sometimes, as a compromise for simplicity and economy, the $\bar{z}(\lambda)$ filter is used for the blue part of $\bar{x}(\lambda)$. Figure 11.2 indicates about the best that can be achieved by this method. Commercial colorimeters often are not this good.

11.2.2 Partial Filtering

In an alternative technique, sometimes referred to as the Dressler principle, some of the filter components are placed side by side so that different portions of the light beam are filtered differently. However, because of technical difficulties in making these rather elaborate stacks of filters, few colorimeters using this principle have been built commercially.

11.2.3 Template Method

This method employs a disperser such as a prism or diffraction grating to produce a spectrum. A template is placed in the spectrum to block different proportions of each wavelength. The light transmitted through the template is re-combined and focussed onto a photodetector. By suitable design of the template, the photodetector gives readings proportional to one of the tristimulus values.

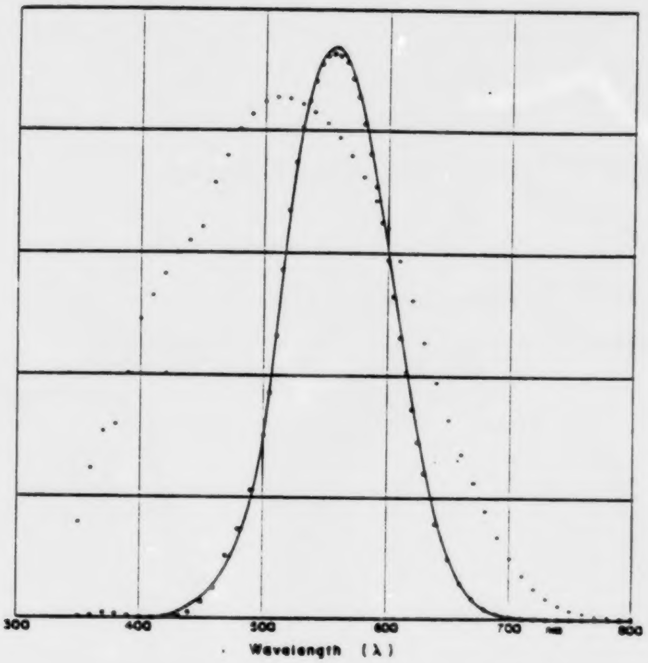

Figure 11.1. Matching of $V(\lambda)$ by full filtering method. (**Budde, 1980**).

11.2.4 Digital Method

In some modern colorimeters a spectrum is focussed on an array of photodiodes. Each photodiode sees only a narrow band of wavelengths. Digital weighting and integration of the signals from the photodiodes gives the three tristimulus values.

Strictly speaking, these instruments are spectrophotometers because full spectral information is available internally. However some are marketed as colorimeters with only the tristimulus values made available to the user.

11.3 Reflection Measurements

For a tristimulus colorimeter to measure reflecting samples it must have its own internal source which should have the spectral power distribution of an appropriate CIE standard illuminant. This can be achieved by filtering techniques similar to those used to modify the detector responsivities.

If the colorimeter is to be used only for non-fluorescent samples (samples that do not change the wavelength of the radiation) the source filter can be combined with the detector filters in a single stack in front of the detector(s).

11.4 Advantages and Disadvantages

The advantages of tristimulus colorimeters over spectrophotometers are that they are faster and less expensive. The disadvantages are that they are less accurate and give less information.

They are well suited to measuring small differences between similar samples because in this case fairly large differences between the instrument's spectral responses and those of the standard observer can be tolerated. Thus tristimulus colorimeters find widespread use in quality control.

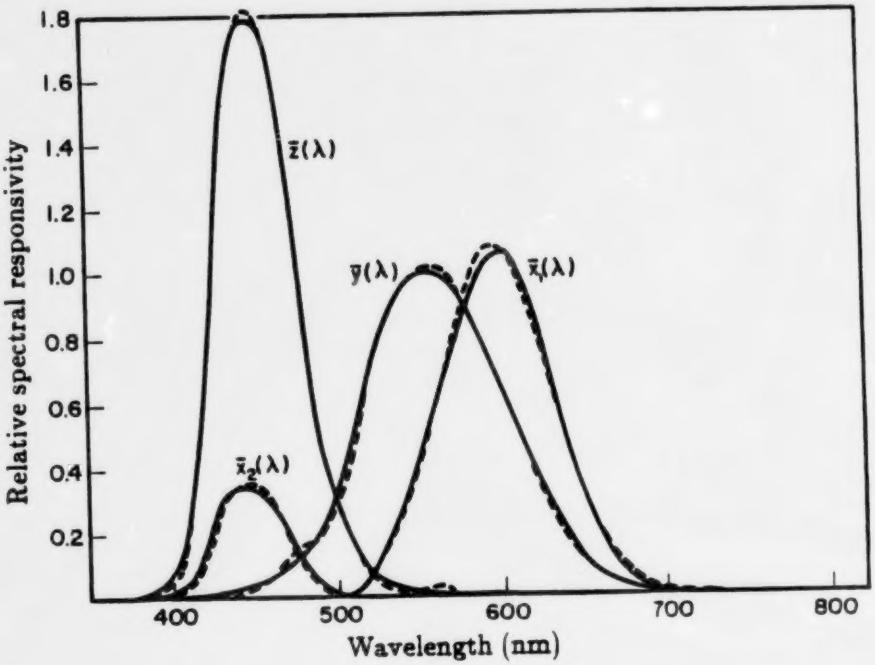

Figure 11.2. Fits to the colour-matching functions in a research colorimeter built at NRC (Robertson et al, 1972). The full lines indicate the colour-matching functions. The dashed lines are the fits achieved.

11.5 References

Budde, W., 1980, **'A Reference Instrument for 20°, 60° and 85° Gloss Measurements'**, *Metrologia*, **16:** 1-5.

Billmeyer, F.W., Jr. and Saltzman, M., 1981, *Principles of Color Technology*, New York: Wiley.

Judd, D.B. and Wyszecki, G., 1975, *Color in Business, Science and Industry*, New York: Wiley.

Robertson, A.R., Staniforth, A., Gignac, D.S., and McDougall, J., 1972, *A Computer-Controlled Photoelectric Tristimulus Colorimeter*, Report No. PRO-387, Ottawa: National Research Council.

Wyszecki, G. and Stiles, W.S., 1982, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, New York: Wiley.

Notes

Lecture 12: Wm B. Cowan

COLORIMETRY OF VIDEO DISPLAYS

The past few decades have seen the growth of a several new media for producing coloured images, the most common of which is the shadowmask CRT (colour television). In recent years these devices have become more and more important as techniques for creating and manipulating images by computer have become more common. As they find applications where exact colour specification is needed we see the colorimetry of these devices becoming very important. Several examples of these applications are:

1. the use of monitors as proofing devices in the printing industry;
2. the use of monitors as design tools in textile design;
3. the use of monitors for colour mixing in the paint industry;
4. and so on. In any of these applications the ability exactly to specify the colour on the monitor is very important because of the need to interchange colour coordinates with other types of colour generation technology. (The easy colorimetric problem addressed in this lecture – how to know the tristimulus values of a monitor colour – is pretty well solved; the more difficult colorimetric problem – how to determine the colour appearance of a monitor colour – is quite unsolved. The solution of the easy problem is, however, a prerequisite for better understanding of the difficult one. Some discussion of the small amount known about calibrating colour appearance is found in Lecture 16: Colour Reproduction.) To begin this subject we will review the device capabilities that must be considered before questions about calibration can sensibly be raised.

Temporal stability

If the device is stable over time it is possible to calibrate it at one time, then use the calibration to determine the tristimulus values of a colour produced at another time. Otherwise put, the precision with which a device is stable over time is the precision to which a calibration done at one time can be used at another time.

Spatial uniformity

Devices for making coloured images produce colour at a variety of locations. If the device is spatially uniform it is possible to calibrate it at one location, then use the calibration to determine the tristimulus values of a colour produced at another location. Otherwise put, the precision with which a device is stable over space is the precision to which a calibration done at one location can be used at another location.

Componential stability

It is often possible to resolve the operation of a complicated colour-generating device into the combined operation of a set of components. If the components are stable and independent of one another in their operation it is possible considerably to simplify a colorimetric calibration by calibrating the components, and not the whole device. The precision to which such components are stable and independent is the precision to which a device can be calibrated by componential analysis.

We will discuss below three different methods for monitor calibration. Which to use depends on the precision with which the monitor can pass these tests compared to the precision to which the calibration is needed. As a preliminary to that discussion there will be a sort description of how colour is produced by a shadowmask CRT, so as to give a little background for the discussion of calibration.

12.1 Light Generation in Self-Luminous Displays

12.1.1 The shadowmask CRT

The shadowmask CRT, the most common colour television technology, consists of the following:

1. A glass envelope, the purpose of which is to maintain a vacuum inside the CRT.
2. Three electron guns, which create beams of electrons directed toward the screen. Each electron gun consists of:
    1. A heated cathode, which is the source of electrons for the electron beam, and
    2. A control grid, which regulates the current in the electron beam. (In the traditional CRT design the grid voltage was regulated to control the beam current; in many modern designs the grid is held constant and the cathode voltage is modulated to control the beam current.)
3. An acceleration grid, which accelerates the electron beams toward the screen.
4. A shadowmask, which confines the beam area of each gun to the appropriate set of phosphor dots.
5. Dots of phosphor, red, green and blue, which are painted on the inside of the glass envelope of the tube. The phosphor is excited by the electron beams, one colour by each gun; when excited it emits light of a characteristic spectral power distribution.

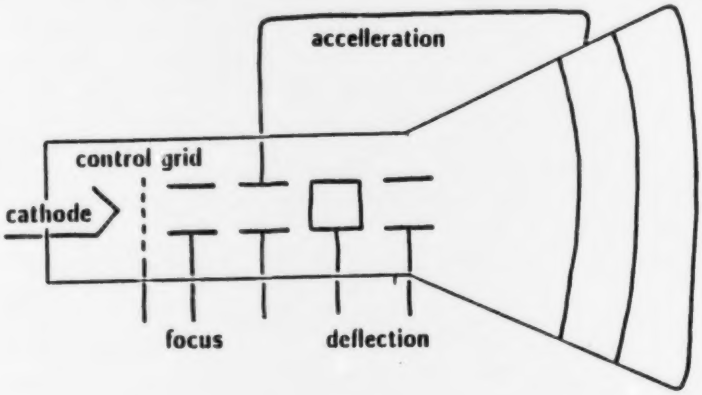

Figure 12.1. Components of the shadowmask CRT.

Taken together these components allow the CRT input to control the relative amounts of three colour components, and thus to control the colour produced. The phosphor dots are very small, as small as 0.2 mm, and close together. The three colours are intermixed so that they combine additively into a single colour when perceived from a normal viewing distance. Thus, the most important colorimetric feature of a CRT is that its colour is the additive mixture of three primaries. Several effects disturb this simple model of a CRT:

1. impurity, where electrons from a single gun strike more than one phosphor dot type;
2. misconvergence, where the red, green and blue images do not exactly coincide on the screen;
3. reflected light, which is light from the room reflected back to the observer from the screen of the CRT; and
4. the black level, which is a uniform level of light emitted by the CRT even when the input is zero.

The first two conditions are caused by monitor misadjustment, and should be eliminated. The third cannot be avoided. The fourth is added to bring the image up to the level of brightness demanded by the light reflected from the screen. Both the third and the fourth can be modelled as a uniform veiling light mixed additively with every colour on the screen.

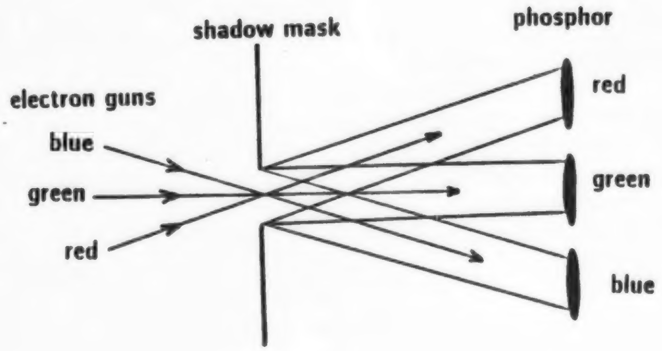

Figure 12.2. Geometrical arrangement of the shadowmask holes and phosphor dots.

12.2 *In Situ* Calibration

The first style of calibration available for monitors takes some of the light from the screen while the monitor is in use. The light is diverted to one or more photosensors which provide tristimulus value output. This provides the ability to know the tristimulus values of a colour on the screen. Producing a colour of known tristimulus values is accomplished using feedback, adjusting the screen colour until the output of the photosensors is correct. We can break down the operation of such a calibration device into several components: light diversion, photosensor, and feedback algorithms, all of which are discussed below.

Light diversion

There are two methods for diverting light from the screen to a photosensor:

1. Use a beamsplitting plate and divert the whole image to a sensor which is able to roam around the image.
2. Place the photosensor in front of a corner of the screen which is not needed for display of the image.

The first produces sensitivity problems. There is little enough light emitted from the screen of a CRT to make colorimetry difficult under the best of conditions. The second requires good predictable spatial uniformity, which is usually the hardest part of monitor performance to manage.

Photosensors

There are several design features that must be resolved when deciding what kind of photosensor to use for an *in situ* calibration system. The first is how well the photosensor resolves the spectrum:

1. At least three photosensors are needed. They are sufficient if there is no light reflected from the screen and if the black level is zero. If not, the readings they give must be properly zeroed, which depends on knowing the monitor stability.
2. The more photosensors are used the lower the sensitivity of the system.

The second is photosensor sensitivity:

1. The lower the sensitivities of the photosensors the longer a reading must be for a given precision.
2. The sensitivity must be good at short wavelength (from 400 nm to 500 nm) to do a good job with the blue phosphor.

The third is the area of capture of the photosensor:

1. A small area of capture allows us to determine the colorimetry of a small area of the screen but severly limits sensitivity.
2. A large area of capture improves sensitivity.

The fourth is the spatial mobility of the photosensor:

1. A photosensor fixed to one area can calibrate only colours which cover that area.
2. A mobile photosensor can examine many areas but is much more complex.
3. A mobile photosensor makes sense only if the photosensor has a small area of capture.

The fifth is the integration ability of the photosensor:

1. Light signals from CRT's consist of a short, intense pulse followed by a long dark period. The photosensor should be able to handle the pulse without saturating and should be able to put out a signal which integrates the amount of light over a fairly long time (16 to several hundred milliseconds).
2. If the photosensor cannot integrate it will be necessary carefully to synchronise the photosensor readout with the scanning mechanism of the CRT.

The sixth is the number of photosensor arrays provided:

1. If one array is provided it can monitor only one area of the screen at a time. Images must be relatively unchanging.
2. Several arrays can calibrate several parts of the screen simultaneously.

Feedback algorithms

The feedback algorithms depend largely on the reading time for the photosensor, since they go through a set, read, calculate, set, read, calculate, set, read, etc. loop. The read time is at least the frame rate of the monitor (16 milliseconds or more). The number of frames required for the read cycle is inversely proportional to the sensitivity of the photosensor.

Advantages

In situ calibration methods have only one advantage, but it is a very big one: they can be used successfully with devices that have any type of instability. The device must only be stable on the time scale that a reading is taken, which is the order of milliseconds.

Disadvantages

There are lots of disadvantages:

1. There must be calibration apparatus between the viewer and the screen at all times.
2. Each calibrated monitor requires a calibration apparatus.
3. There is no practical method for calibrating many areas simultaneously.
4. Photosensors of adequate sensitivity and dynamic range are not available.
5. Feedback algorithms are bound to be slow. This means that although knowing the tristimulus values of an area of the screen may be very fast, making an area that has preselected tristimulus values is likely not to be easy.
6. It is almost impossible to derive the changing tristimulus values of a quickly changing colour.

12.3 Exhaustive Calibration

The second approach to calibration assumes that the monitor has temporal stability at least as precise as needed for the calibration. Then it is possible to calibrate the monitor at one time and use the measurements to calculate trsitimulus values at another time. Here are a few of the issues involved in this approach:

Measurement technique

The basic part of exhaustive calibration is displaying each of the colours the device can produce and measuring its tristimulus values. In the measurement of tristimulus values there are a few things to remember:

1. Many component colours have narrow peaks in their spectral power distributions and/or significant spectral power near the ends of the visible spectrum. For this reason tristimulus colorimetry is usually a bad measurement method and spectra should be taken with a small sampling interval (the order of 2 nm) and corresponding resolution. Work in progress at NRC (Cowan and Rowell, 1986b) should produce more quantitative advice about this point.
2. Light levels are usually fairly low, so that conventional array spectroradiometers need long integration times to get reasonable signal to noise. Carefully designed array spectroradiometers may be able to overcome this limitation (Cowan and Rowell, 1986b).
3. Smoothing of the data may well be useful as a method for dealing with poor signal to noise.
4. For scanning devices like the CRT it is important that the beginning and end of the measurement be synchronised with the scan.

Sampling density

For a system in which only a few colours (up to about 64) will be displayed a complete exhaustive measurement is appropriate. However, system displaying over a million colours are now common. At typical times of 10 to 100 seconds per set of tristimulus values exhaustive measurements take hundreds of days, and are clearly impractical. This problem is typically solved by measuring a subset of the available colours and using an interpolation function (e.g. Saunders, 1986 or Stone et al., 1986). Measurements schemes used in different applications have used 512 to 4096 samples and linear interpolation, but there is no systematic research which shows that these sampling densities are optimal. There are also unresolved questions about how the samples ought to be spread around the colour gamut of the device.

Spatial stability

For devices that have poor spatial uniformity it is potentially necessary to make the measurements at every different point on the screen. In practice this is usually not necessary. Take the shadowmask CRT as a typical example. This device has poor spatial uniformity, with the screen brightness often falling of by as much as 50% from the center to the corners of the screen (measured from a constant viewing position directly in front of the center of the screen). But, CRT designers, aware of the problem, have arranged this effect such that the luminance of the screen declines while the chromaticity remains constant. It turns out that the human visual system requires considerably more spatial uniformity in chromaticity than it does in luminance so that the fall-off in luminance is not important as long as the CRT is adjusted to keep chromaticity constant. Thus, even though the shadowmask CRT seems to have insufficient spatial uniformity for a single calibration to be adequate, in practice the errors made by an inappropriate assumption of spatial uniformity are not serious.

Inverse functions

Sampling and interpolation offers a method for deriving the tristimulus values of a colour produced by a self-luminous device. An inverse function is needed to find the device coordinates needed to produce a stimulus of given tristimulus values. Producing such a function is not simple, and there are many possible solutions. Most have problems when they are asked for the device coordinates of a colour that is outside the gamut of the device. Stone et al. (1986) gives one successful method for producing an inverse function. The method is, however, computationally inefficient compared to the demands of real-time image usage. Thus, more research is needed in this area.

Advantages

There are several advantages to this method of calibration:

1. It can be used for any kind of colour-generating device, not just self-luminous devices. Thus, programs that implement or use the method do not have to be changed when the device is changed.
2. It can be used for devices that are difficult to model, or that fail to which assumptions needed for the modelling process fail to apply.
3. It can be used with spectroradiometers that are too insensitive to allow *in situ* calibration.
4. The measurement process can be automated, so that even when measurement is extensive it can be done without tying up manpower.
5. The forward transform is very efficient to implement.
6. It applies, with qualifications noted above, to any pixel on the screen, so that it is straightforward to build up a complete calibrated image.

Disadvantages

Of course, there are disadvantages:

1. The most significant is parameter storage for the calculations. At a minimum 1,500 floating point numbers must be available for the calculations, and the number may be ten times as large. This problem becomes annually less important as the cost of computer memory declines.
2. The measurements can be quite significant, and may require special equipment to be purchased or built, since the large amount of measurement is reasonable only if it is automated.
3. The problem of constructing an inverse function is not yet satisfactorily solved.
4. The number and spacing of measured samples within the device gamut is not yet resolved.
5. Altering or adjusting the device produces the need for a new calibration.

12.4 Model-Dependent Calibration

The final approach to monitor calibration also assumes that the monitor has temporal stability. It also assumes that there is a simple model that applies to its operation, that the monitor can be adequately represented as a collection of independent non-interacting components. Thus, assumptions must also be made concerning the independence of the components. As an example we give a short sketch of the dominant model for shadowmask CRT's. It is followed by a discussion of the advantages and disadvantages of model-dependent calibration in general.

The additive mixture model for shadowmask CRT's

The additive mixture model assumes that the colours produced by a CRT is an additive mixture of the colours produced by each of the three phosphors, and that the colour produced by each phosphor is independent of the colour produced by each of the other phosphors. It is generally further assumed that the colour produced by each phosphor has a chromaticity that is independent of the level at which the phosphor is excited. These two assumptions are known as 'gun independence' and 'phosphor constancy'. How to measure them for a particular monitor is described in Cowan and Rowell (1986a). These are the additional assumptions that must be true if a model-dependent calibration based on the additive mixture model is to be used as the basis of a calibration. Knowing that the assumptions are true it is then possible to use a simplified model of the colour produced by each phosphor. It depends on:

1. the chromaticity of the light emitted by the phosphor,
2. the level of light emitted in response to a given input to the CRT,
3. the relative excitations of the various phosphors, and
4. the overall brightness of the CRT.

The usual procedure is to measure some aspects of CRT function, then to calculate parameters that describe the components of the phosphor model using the measurements. When the tristimulus values of a particular colour are needed the parameters are used in conjunction with calibration equations to determine them. When a colour of specified tristimulus values is needed the parameters are used in conjunction with the inverse functions to calculate the device coordinates needed to produce that colour. The details of the functions needed and of the measurements needed to determine the parameters of a particular CRT are to be found in Cowan (1983) or Cowan (1984).

Advantages

Here is a list of the advantages of this type of calibration scheme:

1. If the model is chosen carefully the forward function and its inverse are equally easy to implement.
2. There is usually less measurement needed for this type of calibration than for any other type.
3. There are usually fewer parameters to store than for any other type of calibration.
4. The calibration scheme will often not work properly if the monitor is seriously out of adjustment. This may be either an advantage or a disadvantage.
5. Calibration schemes exist that require only radiometry, and not spectroradiometry.
6. Subject to the qualifications discussed in section 12.3, the calibration applies to every pixel on the device.

Disadvantages

There are also significant disadvantages:

1. The forward function is usually more expensive than for exhaustive calibration methods.
2. The validity of the model chosen must be tested before calibrating a device.
3. While a model will apply with different examples of one device it is very unlikely to apply to more than one device type. Thus new calibration experiments, measurements routines and functions must be found for each new device type.
4. Altering or adjusting the device produces the need for a new calibration.

12.5 References

Cowan, W. B., 1983, 'An Inexpensive Scheme for Calibration of a Colour Monitor in terms of CIE Standard Coordinates', *Computer Graphics*, **17**: 315–322.

Cowan, W. B., 1984, 'CIE Calibration of Video Monitors', National Research Council of Canada Technical Report, Number 25362.

Cowan, W. B. and Rowell, N. L., 1986a, 'On the Gun Independence and Phosphor Constancy of Colour Video Monitors', *Color Research and Application*, **11**: S34–S38.

Cowan, W. B. and Rowell, N. L., 1986b, work in progress at NRC.

Saunders, B., 1986, 'Implementation of Color Transforms in Electronic Imaging Systems', submitted to *Journal of Imaging Technology*.

Stone, M., Cowan, W. B., and Beatty, J. C., 1986, 'Color Gamut Mapping and the Printing of Digital Color Images,' in preparation for *ACM Transactions on Graphics*.

Notes

Lecture 13: A. R. Robertson

GLOSS MEASUREMENTS

13.1 Introduction

Gloss is an attribute of a surface by which reflections of a light source are seen as highlights superimposed on the diffusely reflected light from the surface. Along with other attributes such as haze, lustre, translucency, sheen and texture, it is associated with the geometric distribution of reflected light.

This lecture will be restricted to a discussion of standard methods of measuring specular gloss. Specular gloss is gloss associated with reflection that is regular (ie. specular or mirror-like) on a macroscopic scale.

Good surveys of gloss measurement techniques can be found in CIE (1986), Hunter (1975) and Judd and Wyszecki (1975).

13.2 Standard Methods for Specular Gloss Measurement

The luminance of a surface as a function of viewing direction can be illustrated by goniophotometric curves such as those in figure 13.1.

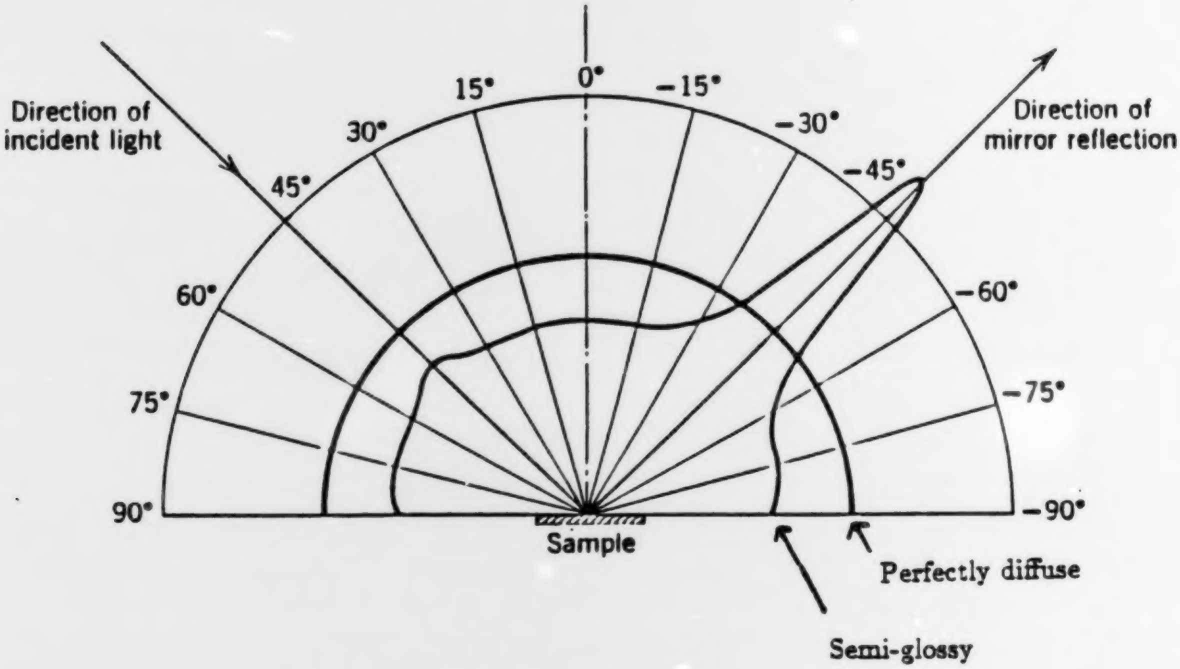
Direction of incident light
Direction of mirror reflection
0°
15°
30°
45°
60°
75°
90°
−15°
−30°
−45°
−60°
−75°
−90°
Sample
Perfectly diffuse
Semi-glossy

Figure 13.1. Goniophotometric curves for a perfectly diffuse and a semi-glossy surface when illuminated at 45° (Judd and Wyszecki, 1975).

A perfectly diffuse (perfectly mat) surface has the same luminance in all directions and thus is denoted by a semicircular curve. Progressively more glossy surfaces have progressively narrower peaks of reflection in the specular direction (at the same angle to the normal as the incident beam but on the other side of the normal).

Standard methods of measuring specular gloss involve arrangements such as is illustrated in figure 13.2. The angles of incidence and viewing, and the sizes of the source and receiver apertures are defined rigidly. The amount of light reflected through the receiver aperture is a measure of the specular gloss.

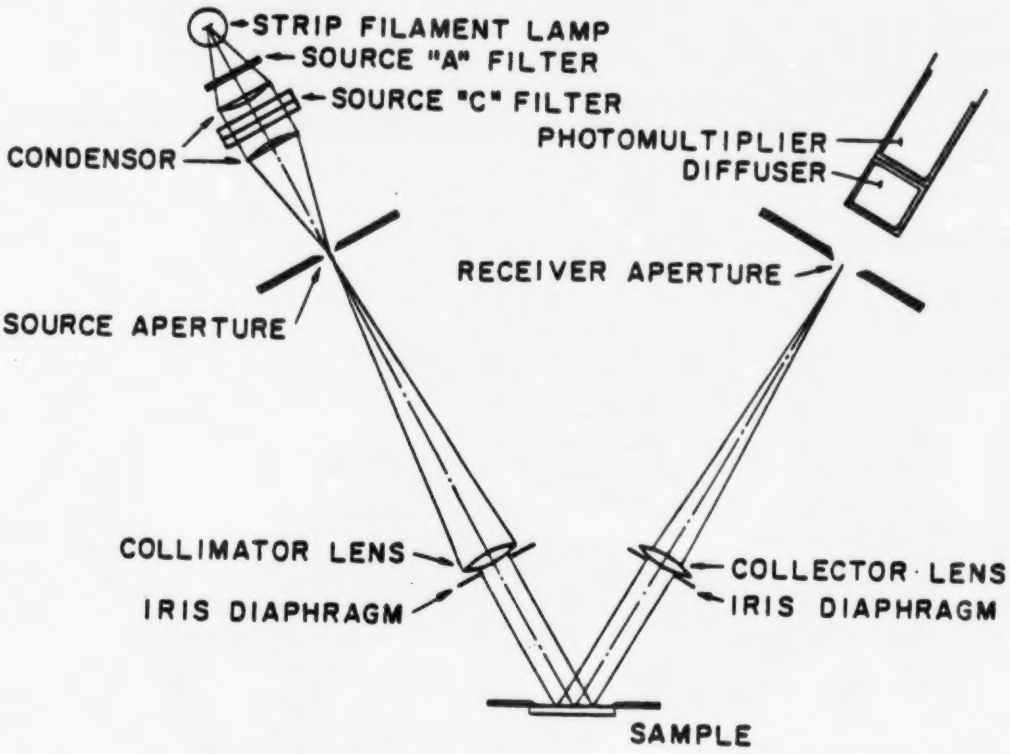

Figure 13.2. Schematic diagram of NRC reference glossmeter (Budde, 1980a).

The best angle for the measurement depends on the type of material, with large angles being best for low-gloss materials and small angles for high-gloss materials. The most commonly used angle is 60°. Angles that have been used for different materials are shown in table 13.1.

Various national and international standards have been written to define the angles and aperture sizes to be used in specular gloss measurements. Some of these are listed in table 13.2

13.3 NRC Reference Glossmeter

NRC operates a reference glossmeter that has been designed for accurate methods according to ISO Standard 2813 or ASTM D523, which is equivalent. The instrument is illustrated in figure 13.2.

In most instances the deviations or uncertainties of the geometrical parameters are considerably smaller than the permitted tolerances. The spectral conditions are better than in most commercial glossmeters and polarization effects have been completely eliminated. The uncertainties of measurements with the instrument are less than 0.3 gloss units.

Industrial Application	85° sheen	75° specular	60° specular	40° specular	20° specular	0° specular
Paints	Matte paints		Medium-gloss paints		High-gloss paints	
Plastics			Medium-gloss plastics	Clear plastic	High-gloss films and laminates	
Paper		Coated papers, inks			Cast-coated papers, waxed paper inks	
Ceramics			Medium- and low-gloss tiles	Abrasion of all tiles, porcelains		
Textile						
Wax and Finishing			Waxes, floor finishes			
Automotive					Automotive finishes	
Appliance					Finishes	Enamels
Metals						Metal surfaces
General	Low- to medium-gloss materials	Low- to medium-gloss materials	Medium-gloss materials	Medium-to high-gloss materials	High-gloss materials	High-gloss materials

Table 13.1. Specular gloss methods of various industries (CIE, 1986).

13.4 Reference Standards

The ultimate reference or primary standard for gloss measurements is a glass surface having a refractive index of 1.567. This surface has, by definition, a gloss value of 100 at every angle. (Some details and limitations of this definition are discussed by Budde, 1980b).

A material with exactly such a refractive index will rarely be found but practical standards can be produced by measuring the refractive index of a suitable material and calculating its reflectance relative to the primary standard for the defined angles.

The most commonly used standard is a piece of polished black glass. However the refractive index, and consequently the gloss value of such a standard has been found to decrease. The change can amount to about 1 gloss unit in a period of 3 to 5 years. Because of this a more stable standard has been developed at NRC. This is in the form of a fused-quartz wedge as illustrated in figure 13.3. The standard is wedge-shaped to deflect the reflection from the back surface away from the detector optics.

This quartz-wedge standard cannot be used in some commercial glossmeters because the back-surface reflection can enter the detector optics. It is most suited for the checking and calibration of black glass standards that can be used in the field.

13.5 Conclusion

Specular gloss is measured by measuring the amount of light reflected under carefully controlled geometrical conditions. A quartz-wedge reference standard can be used to calibrate black-glass working standards.

Industry	Specification	Angle of incidence 20°	45°	60°	75°	85°
Paint	ISO 2813	A		C		E
	(UK) BSI 3900 Part D2	A		C		E
	(FRG) DIN 6753	A		C		E
	(USA) ASTM D523	A		C		E
Plastics	(USA) ASTM D2457	A	B	C		
Wax, Polish	(USA) ASTM D1455			C		
Ceramic	(USA)ASTM C346		B			
Paper	ISO DIS ...				D	
	(USA) ASTM D1223				D	

Geometry	Source Aperture (°) in plane of measurement	Source Aperture (°) normal to plane of measurement	Receptor Aperture (°) in plane of measurement	Receptor Aperture (°) normal to plane of measurement
A	0.75	3.0	1.8	3.6
B	1.4	3.0	8.0	10.0
C	0.75	3.0	4.4	11.7
D	2.8	5.7	11.4 diameter	
E	0.75	3.0	4.0	6.0

Table 13.2. Geometries of some standard gloss measurement procedures.

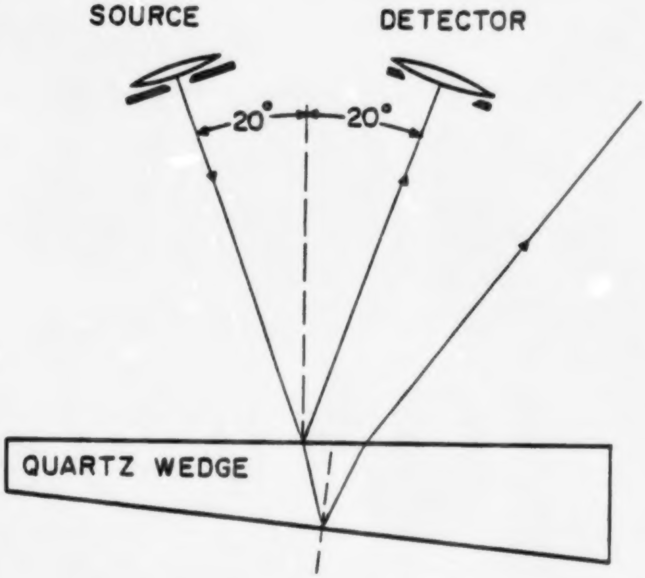

Figure 13.3. Quartz-wedge gloss standard (Budde, 1980a).

13.6 References

Budde, W., 1980a, 'A Reference Instrument for 20°, 60° and 85° Gloss Measurements'. *Metrologia*. **16**: 1-5.

Budde, W., 1980b, 'The Calibration of Gloss Reference Standards'. *Metrologia*. **16**: 89-93.

CIE, 1986, *Evaluation of the Attribute of Appearance Called Gloss*, Paris: Commission Internationale de l'Eclairage.

Hunter, R.S., 1975, *The Measurement of Appearance*, New York: Wiley.

Judd, D.B. and Wyszecki, G., 1975, *Color in Business, Science and Industry*, New York: Wiley.

Notes

Lecture 14: D. E. Minns

TECHNOLOGY TRANSFER

14.1 How to Access the NRC Laboratories

NRC has representatives across Canada and you can:

1. Find the telephone number of your local contact in the yellow pages under 'Technology'.
2. Contact the Laboratories, directly perhaps through somebody you know who will direct you to the Industrial Liaison Officer, who for Division of Physics is
   Dr. D. E. Minns (613) 993-3235.
3. Call the main NRC telephone number
   (613) 993-9101.

14.2 Getting Started

If you have a commercial opportunity requiring technical input from the NRC Laboratories the industrial liaison officer is a key person for you to know and get involved in the project.

He/She will try to put together with you a plan tailored to the particular opportunity combining

- Manpower Resources
- Financial Resources
- Schedules
- Technology

14.3 Resources

Generally any given Laboratory has limited manpower resources and zero financial resources to apply to your opportunity. However, the Industrial Liaison Officer will know where these are available and how to bring them to bear on the particular opportunity.

14.3.1 Sources of Manpower

1. The Laboratory: limited but some effort will always be made available to assist in directing a commercial development program and in technology transfer.
2. The Company: This is an important source since ultimately the company must take the responsibility for the success of the commercial venture and must actively participate in the technical development and transfer process.
3. Other: Where 1. and 2. are both limited or the project's time schedule demands extra assistance - then technical personnel can possibly be brought in under contract.

14.3.2 Sources of Finance

1. The NRC Laboratory: has no finances to assist in commercial exploitation - appropriate sources are:
2. Other Government Agencies:
    - NRC Industry Development Office (IRAP Program)
    - DRIE
    - Others (e.g. DSS)
3. The Company: Since it is the company who stands to gain monetarily from the enterprise, substantial commitment will be expected financially from the company usually in the form of equipment purchase, supplies and materials, staff overheads etc.
4. Venture Capitalists: NRC has frequent visits from such groups and may be able to assist in identifying sources.

14.3.3 Sources of Technology

1. Technology needs developing: A collaborative R and D program will be set up in partnership with the company generally involving company staff conducting work at NRC facilities under NRC direction although it could be an NRC staff secondment to the company. Existing NRC equipment and facilities will generally be made freely available but any additional equipment or costs, manpower etc will have to be sought elsewhere (see Sources of Finance and Manpower).
2. Technology exists at NRC: The company will be licensed to use existing technology.
3. Technology exists off-shore: NRC can assist the company in assessing the technology and can help the company adopt the technology. The company must seek permission to use the technology.
4. Technology exists elsewhere in Canada: See 3 above.

14.4 Conclusion

We are very keen to help Canadian industry keep pace with changing technologies to maintain or improve its competitive edge. It is largely industry's responsibility to identify commercial opportunities, and it is largely NRC's responsibility to make sure that it has relevant expertise, facilities and technology. The Industrial Liaison Officer of the Laboratory Divisions of NRC is responsible for combining these resources in a way that can take maximum advantage of these opportunities.

Notes

Notes

Lecture 15: A. R. Robertson

COLOUR-DIFFERENCE EVALUATION

15.1 Introduction

Numerical specification of colour differences, based on instrumental measurements, is important in many applications. Often one needs to know how close a manufactured product is to the required colour, how well two samples of the same product match each other, or how large are the errors in a colour image reproduction process. This lecture will describe formulae by which indices of colour difference can be calculated. More detail can be found in the usual textbooks (Billmeyer and Saltzman (1975), Judd and Wyszecki (1975), and Wyszecki and Stiles (1982)).

15.2 Visual Evaluation

To evaluate the difference of colour between two samples it is necessary to view the samples side-by-side in a suitable level and quality of illumination such as is provided in a viewing booth. Most precise evaluation is obtained with high illuminance (1000 lux), fine dividing line, large area and binocular viewing.

In many circumstances, though, visual evaluation is not convenient and it is desirable to measure the tristimulus values of each colour and to calculate an index of colour difference (usually denoted by ΔE) that represents the perceived size of the difference between the two colours.

15.3 Colour Differences in X, Y, Z or Y, x, y Space

The simple Euclidean distance in X, Y, Z or Y, x, y space is not a suitable index of colour difference. Equal distances do not by any means represent equally perceptible differences. Figure 15.1 illustrates this in the x, y chromaticity diagram.

Each ellipse represents the locus of colours equally different from a test colour at the centre of the ellipse. If the diagram gave a uniform representation of colour differences the ellipses would all be circles of equal radius.

15.4 Historical Formulae

Many formulae have been suggested to calculate colour differences from pairs of X, Y, Z values. A list of some of the formulae in use in 1976 is given in table 15.1.

15.5 Uniform Colour Spaces

Many of these formulae are based on a uniform colour space. A uniform colour space is formed by transforming X, Y, Z (or Y, x, y) into three new quantities that are plotted in rectangular coordinates. The new quantities are defined in such a way that equal distances in the uniform colour space correspond much more closely to equally perceptible colour differences than do distances in X, Y, Z or Y, x, y space.

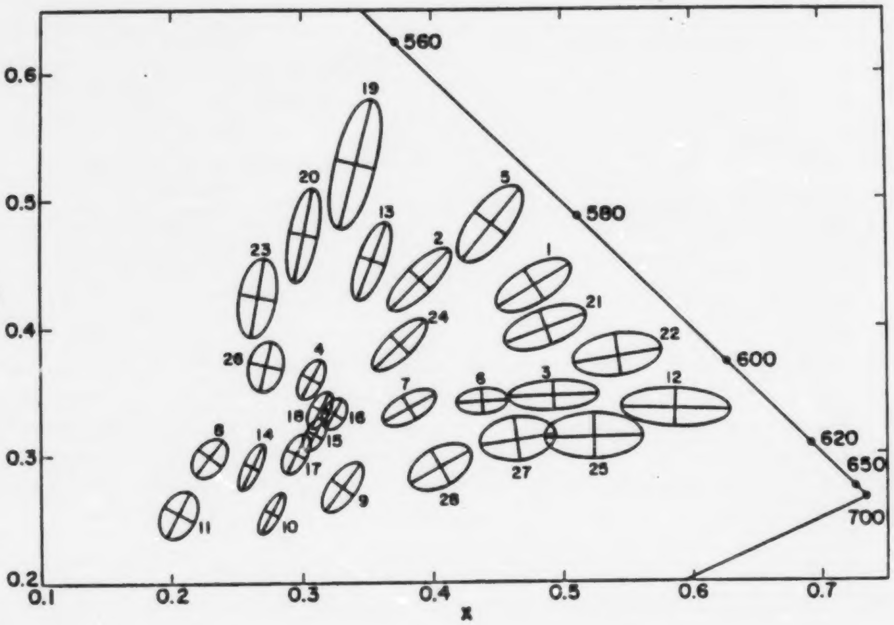

Figure 15.1. Colour matching ellipses in the CIE 1931 chromaticity diagram (Wyszecki and Fielder, 1971).

Adams Chromatic Valence
Adams-Nickerson
CIE 1964
DIN
Friele-MacAdam-Chickering 1
Friele-MacAdam-Chickering 2
Glasser et al cube-root
Godlove-Munsell
Hunter Lab
Judd-Hunter ('NBS Unit')
MacAdam/Simon-Goodwin
Saunderson-Milner
Scofield

Table 15.1. Examples of colour-difference formulae in use in 1976.

15.6 CIELAB and CIELUV Colour Spaces

15.6.1 CIE 1976 Recommendations

The use of so many different formulae was confusing, especially as it was not possible to convert ΔE values calculated by one formula to those calculated by another. Thus in 1976, the CIE recommended the use of two approximately uniform colour spaces and associated colour-difference formulae (Robertson. 1977). The spaces and formulae were chosen from amongst several of similar merit to promote uniformity of practice, pending the development of a space and formula giving substantially better correlation with visual judgements.

Experimental evidence then available was considered insufficient to warrant the selection, on merit alone, of a single color-difference formula which would be satisfactory for most industrial applications. However, the adoption of the two spaces and formulae of similar merit but different structure was considered to go a long way toward unification of current practice.

One of these uniform colour spaces, known as the CIE 1976 (L^*, a^*, b^*) space (abbreviated CIELAB), is produced by plotting in rectangular coordinates the quantities L^*, a^*, b^* defined by:

$$L^* = 116(Y/Y_n)^{1/3} - 16$$
$$a^* = 500\left((X/X_n)^{1/3} - (Y/Y_n)^{1/3}\right)$$
$$b^* = 200\left((Y/Y_n)^{1/3} - (Z/Z_n)^{1/3}\right)$$

where X_n, Y_n, Z_n define the colour of the nominally white object-colour stimulus. Usually this is one of the standard illuminants, A or D_{65}, with Y_n equal to 100. For very low tristimulus values, the following changes must be made:

if $X/X_n < 0.008856$, then $(X/X_n)^{1/3}$ becomes $7.787 \cdot X/X_n + 16/116$

if $Y/Y_n < 0.008856$, then $(Y/Y_n)^{1/3}$ becomes $7.787 \cdot Y/Y_n + 16/116$

if $Z/Z_n < 0.008856$, then $(Z/Z_n)^{1/3}$ becomes $7.787 \cdot Z/Z_n + 16/116$

The total colour difference between two colours that differ by L^*, a^*, b^* is given by:

$$\Delta E^*_{ab} = \left((\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2\right)^{1/2}$$

The other uniform colour space is the CIE 1976 (L^*, u^*, v^*) space (abbreviated CIELUV). It is produced by plotting in rectangular coordinates the quantities L^*, u^*, v^* defined by:

$$L^* = 116(Y/Y_n)^{1/3} - 16$$
$$u^* = 13L^*(u' - u'_n)$$
$$v^* = 13L^*(v' - v'_n)$$

with

$$u' = 4X/(X + 15Y + 3Z)$$
$$v' = 9Y/(X + 15Y + 3Z)$$

where Y_n, u'_n, v'_n define the colour of the nominally white object colour stimulus. The quantity L^* is the same as in CIELAB and has the same modification when $(Y/Y_n) < 0.008856$. The total colour difference in this system is

$$\Delta E^*_{uv} = \left((\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2\right)^{1/2}$$

The merits of CIELAB and CIELUV as uniform colour spaces and colour-difference formulae are about equal. Their preference in different industries depends on other factors. The CIELUV system is often preferred in industries dealing with additive colour mixture (television, video displays etc.) because it incorporates a chromaticity diagram (the u', v' diagram) in which additive mixtures plot on straight lines. The CIELAB system is preferred in colorant-producing industries (dyes, paints, pigments etc.) mainly because of its similarity to the older Adams-Nickerson formula.

The rest of this lecture will refer only to the CIELAB system, although most of what is said would apply equally well to the CIELUV system.

15.6.2 Tolerances

Under ideal conditions a difference of 0.2 to 0.5 CIELAB units can be detected although a commercial match is normally about 1.0 units. These numbers are approximate because CIELAB space is not perfectly uniform and the size of a just perceptible difference varies a little from one part of colour space to another. Because of this it is quite common to set different ΔE^*_{ab} tolerances for each colour in which a particular product is produced.

15.6.3 L^*, a^*, b^* Plots

If the values of L^*, a^*, b^* of samples of a particular product are plotted, it is often possible to draw lines that enclose all the acceptable matches. This type of presentation has two advantages. Firstly it allows different tolerances to be applied to each of the three axes. Secondly it gives an indication of the nature of a colour difference as indicated in table 15.2

	+	-
L^*	Lighter	Darker
a^*	Redder	Greener
b^*	Yellower	Bluer

Table 15.2. Interpretation of differences in L^*, a^*, b^*.

15.6.4 CIELAB Chroma and Hue

Many people find it easier to interpret colour differences in terms of differences of hue and chroma rather than redness, greenness, blueness and yellowness as in table 15.2. The CIELAB system provides for this by allowing correlates of hue and chroma to be calculated as follows:

$$h_{ab} = \tan^{-1}(b^*/a^*)$$

$$C^*_{ab} = \left((a^*)^2 + (b^*)^2\right)^{1/2}$$

The quantity h_{ab} is known as CIELAB hue-angle, and the quantity C^*_{ab} is known as CIELAB chroma. Similarly the quantity L^* is known as CIELAB lightness because it correlates well with perceived lightness.

The correlation of CIELAB hue-angle and chroma with the corresponding perceived attributes is not perfect but there is a general correspondence in the sense that changes of CIELAB hue-angle correspond to changes of perceived hue and changes of CIELAB chroma correspond to changes of perceived chroma.

For ease in comparing the sizes of lightness, chroma and hue differences, CIELAB hue difference is usually specified by ΔH^*_{ab} defined by:

$$\Delta H^*_{ab} = \left((\Delta E^*_{ab})^2 - (\Delta L^*)^2 - (\Delta C^*_{ab})^2\right)^{1/2}$$

The total colour difference, ΔE^*_{ab} is then the quadratic sum of three component differences:

$$\Delta E^*_{ab} = \left((\Delta L^*)^2 + (\Delta C^*_{ab})^2 + (\Delta H^*_{ab})^2\right)^{1/2}$$

For small colour differences away from the achromatic axis,

$$\Delta H^*_{ab} = C^* \Delta h^*_{ab}(\pi/180)$$

Using these new coordinates, colour tolerances can, if needed, be set as separate ΔL^*, ΔC^* and ΔH^* values for each product. This technique has proved to be very useful, both in setting tolerances and in describing colour differences.

The dyeing industry in the United Kingdom has followed this approach and has even developed a complicated empirical formula for determining the tolerances (Clarke et al. 1984).

15.7 Conclusion

Colour differences can be specified by means of the CIELAB formula in terms that are meaningful and understandable in most industries throughout the world. In some industries the alternative CIELUV formula is used. The most complete and useful method is to specify differences of lightness, chroma and hue separately by means of ΔL^*, ΔC^*_{ab} and ΔH^*_{ab}.

15.8 References

Billmeyer, F.W., Jr. and Saltzman, M., 1981, *Principles of Color Technology*, New York: Wiley.

Clarke, F.J.J., McDonald, R., and Rigg, B., 1984, 'Modification to the JPC79 Colour-Difference Formula', *J. Soc. Dyers Colourists*, **100**: 128-132.

Judd, D.B. and Wyszecki, G., 1975, *Color in Business, Science and Industry*, New York: Wiley.

Robertson, A.R., 1977, 'The CIE 1976 Color-Difference Formulae', *Color Res. Appl.*, **2**: 7-11.

Wyszecki, G. and Fielder, G.H., 1971, 'New Color-Matching Ellipses', *J. Opt. Soc. Am.*, **61**: 1135-1152.

Wyszecki, G. and Stiles, W.S., 1982, *Color Science, Concepts and Methods, Quantitative Data and Formulae*, New York: Wiley.

Notes

Lecture 16: Wm B. Cowan

COLOUR IMAGE REPRODUCTION

16.1 Principles of Colour Reproduction

There are special problems which arise when a colour image is reproduced, whether in the same medium (as when a colour CRT image is presented on a different CRT) or in a different medium (as when a colour CRT image is reproduced as a slide or print). In this lecture we discuss colour reproduction as if it is a special case of colour difference. As such the principal problem in colour reproduction becomes the discovery of a special kind of colour difference metric, one that, given two images, describes how much difference an observer perceives between them. The obvious place to start is to consider colours metrics of the type discussed in lecture 15. If we had a metric of that type we could compare the picture point by point with the reproduction and then, by aggregating of the colour differences at each point, come up with a specification for how different in colour the two images are. However, it turns out that this is not a practical solution. Consider the following problems:

1. An attempt to produce the same tristimulus values point for point is, except in a few limiting cases, prohibitively expensive.
2. Reproduction in a different medium means that such factors as surface texture and gloss cannot be reproduced. Differing surface characteristics mean that the colour appearance is not identical even if the tristimulus values are the same.
3. Different media can have different colour gamuts so that exact reproduction may be simply impossible.
4. The correct method for aggregation is certainly not a simple sum, since the amount of mismatch at one point affects the perceived mismatch at a neighbouring point.
5. The correct method of aggregating colour difference is likely to depend on the content of the image.

The truth is that in these cases we are almost never looking for a reproduction which is exact in terms of colour, but one which represents adequately the important features of the original. This is not an easy task, since we do not, in general, know what the important features of an image are. The next section, which discusses some general properties of colour difference perception sheds a little light on what the important features may be, but we have to face the fact that this is an area where science has not even begun to scratch the surface. For a long time this area of colour difference evaluation will remain more of an art than a science.

Visibility of Colour Differences

When we discuss colour reproduction as a special case of colour difference we have to look for those special properties of colour difference that apply to the appearance of colours. (We say colour appearance because our object in colour reproduction is to make the reproduction have as close an appearance to the original as money and the reproduction medium allow.) Here is an short list of properties of colour difference that seem to be relevant to appearance. It is not comprehensive, and, since colour appearance is an area where little is known, it raises more questions than it answers.

Colour difference depends on the mode of perception.

The idea of a colour difference metric is to produce an equation which, given the tristimulus values of two areas of colour, will yield a number which corresponds to the perceived difference in their colours. Consider, however, the following cases:

1. Two areas of colour are shown on two identical CRT's. One is adjusted until the two have exactly the same tristimulus values. Then, the perceived colour difference between the two colours is zero.
2. An area of colour is shown on a CRT, and a similar area of colour, adjusted to have exactly the same tristimulus values, is produced by printing on paper. Most observers will now say that the two areas of colour are not exactly the same, that the difference between them is greater than zero (because of differences in gloss, surface texture, etc).
3. An area of colour printed on paper is given, and the colour of an area on a CRT is adjusted until it is as close as possible to the colour of the paper. Do the two areas have the same tristimulus values? The answer is not known at present, but this is currently an area of active research.

What we can conclude is that the perceived colour difference depends not only on the tristimulus values of two colours, but also on other factors such as gloss, surface texture, and so on. We can also say that we do not know at present the answers to even very elementary questions about this phenomenon.

Colour difference depends on surrounding colours.

Consider an image with two or more colours, and compare it to another image which is the same in every respect except that we have changed some or all of the colours. Let us compare two types of colour change:

1. In the first case we change the colour of one area in the image, and ask, comparing the two images, how large the colour change is.
2. In the second case we change the colour of all the areas in the image in the same direction. That is, if the test area is made more reddish, then all the other areas in the scene are also made more reddish by roughly the same amount. Once again we ask, for the same area as in case one how large the colour change of the test area is, comparing between the two images.
3. In the third case we also change the colour of all the areas in the image. But this time we change the test area in one direction and all the other areas in the other direction. That is, if the test area is made more reddish all the other areas are made more bluish (cyan). Then, for this case, we ask how large the colour change of the test area is.

Intuition tells us that the perceived colour change is largest in the third case, smallest in the second case, and intermediate in the first case. The point is that colour difference is perceived relative to other colours. The visual system more or less automatically looks not at the absolute colours in an image, but at the relationships between colours in the image.

Colour difference depends on the state of adaptation.

If you go from a room where the illumination is white into a room where the illumination is reddish there is a period of time during which everything looks to be reddish in colour. This experience of reddishness gradually wears off, to a point where things in the room with reddish light look pretty well the way they would look if the illumination were to be neutral coloured. The process whereby the visual system alters its properties to take account of the overall stimulation being received is called adaptation, and we say that the visual system is changing its state of adaptation during the time in which the reddish cast is disappearing. You should note that when adaptation changes the appearance of a coloured area changes even though its tristimulus values do not change. You should also note that adaptation is not complete; you can always tell, if you have to, whether the illumination in the room is reddish or neutral in colour, even when adaptation has proceeded as far as it can go. This latter property makes it very difficult to do research on adaptation which produces results which are reproducible across observers.

Under normal adaptation our sensitivity to colour differences is by far the most sensitive when the colours are close to white. That is, a slight pinkish hue, caused by a small change in the tristimulus values in the direction of red, is quite noticeable compared to a white sample. However, if the two samples are saturated red, the same change in tristimulus values, slightly reddening one sample compared to the other will not cause a noticeable colour difference. These effects can be seen in the colour difference formulas (CIELAB and CIELUV) described in lecture 15. It appears that this sensitivity increase occurs near the colour which appears to be white under the prevailing state of adaptation, not near any objective white colour. (Hence the use of a white colour in the CIELAB and CIELUV formulas.) Thus the colour difference between two samples of constant tristimulus values changes as the observer's state of adaptation changes.

Colour difference is relative.

Consider two areas of colour which are close to one another on a black background. They are, at the beginning, exactly the same white or grey colour. We will do one of two things to the areas, either make one of them slightly redder, or make the other one slightly less red, which will make it appear slightly bluish-green (cyan). Looking carefully at the two areas you will be unable to tell which of the two actions has been taken; you will only be able to say that one of the areas is slightly redder than the other. On the other hand, if we were to change the background from black to a grey colour with chromaticity (not necessarily luminance) close the same chromaticity of the original white colour, then it will be easy to tell whether the whether red has been added to one area or subtracted from the other.

This demonstration illustrates two principles of colour difference.

1. Colour difference is relative and applies between two colours. It means essentially the same to say that area A is redder than area B as to say that area B is more cyan than area A. Only when a third reference area (the background) is available to compare each area individually to is it possible to make statements about the colour differences of single areas.
2. It is much easier to make fine colour difference judgments when the two areas are visible simultaneously than it is when the two areas are seen at different times. A similar effect occurs with distance; it is much easier to see colour differences when the two areas are seen close together than when the two areas are far apart.

Colour difference depends on the size of the coloured areas.

Consider two adjacent areas which are close to being the same in colour. If the areas are large (the order of ten degrees of visual angle or larger) then the perceived colour difference between them will be larger than if the two areas are small (the order of two degrees of visual angle or less). This effect is well known intuitively; you get close to the thing you are inspecting if you want to check out the possibility of there being a small colour difference in it. The effect is also the basis of a colorimetrist's joke (told by Robert Boynton):

'When rug was laid in my house the two pieces that were laid down weren't the same colour and now I have an ugly line across the middle of my living room where the rug colour changes. How can I get rid of it?'

'Look at the rug through a mailing tube.'

There is another effect associated with size. The yellow/blue channel of the opponent colours scheme (lecture 3) is thought to integrate over a larger area than either the red/green channel or the luminance channel. Thus colour differences which have a yellow/blue component are more washed out as the stimulus size decreases than colour differences with no yellow/blue component.

Colour difference depends on juxtaposition of colours.

The areas which are being inspected for colour difference can be presented in many different ways, side by side, separated by a background area, one after the other, and so on. How they are presented affects the perceived colour difference. We will just give a few examples, since these effects are not comprehensively understood.

1. If two areas which are to be inspected for colour difference are viewed sequentially then it is only possible to see a small colour difference if there is a standard available which can be compared to both of the areas.
2. The most acute colour difference perception occurs when the two stimuli are viewed simultaneously with a narrow (about 10 minutes of visual angle) black border between them.
3. When two colour stimuli exactly abut one another (no dividing line) sensitivity to yellow/blue colour differences is low.
4. As two colour stimuli are moved farther apart the perceived colour difference between them declines (provided the colour difference is small).

The above features of colour difference apply to all types of colour difference judgment. We have introduced them here because they show how complicated the assessment of colour difference is when the images to be judged become complex. All the effects noted above, and many more not yet discovered, no doubt, occur frequently in natural images. Thus, a colour difference metric that can be used as a guide for colour reproduction must take all of them into account.

Applicability of Colour Difference Metrics

Colour metrics are at issue when we talk about reproduction because of the possibility that we can evaluate the 'goodness' of a reproduction by summing a colour difference formula pixel by pixel. The discussion of colour difference immediately above should make it clear that this is not a realistic proposal. There is indeed something like a colour difference metric associated with colour reproduction problems of the sort discussed here, in the sense that, given an original and a reproduction, it is possible to assign a psychophysically determined number which measures how good the reproduction is. The complete compilation of all possible numbers would constitute a colour difference metric for the reproduction problem. The point of this discussion is threefold:

1. We are far from being able to specify an algorithm which can assign such numbers. Research is badly needed in this area, and is well worth doing.
2. Exhaustive experimentation is not useful as the basis of such a metric; the parameter space is too big. An understanding of colour appearance is what we need as the basis of the metric.
3. Such a metric is unlikely to have a close connection with currently available colour difference formula like CIELAB and CIELUV, which have applications quite distinct from colour reproduction.

16.2 Types of Reproduction

Colour reproduction problems occur whenever a colour which exists in one location or at one time is to be reproduced at another location or time. This problem has been a significant issue in the photographic, printing and other industries for a long time. Each industry has worked out rules of thumb which allow it to produce reasonable colour reproduction on a routine basis. For the most part these rules of thumb are based on detailed intuition about the reproduction processes, and not on any type of colour difference formula. The one situation where such formulas are in frequent use is where large areas of a single colour are to be reproduced. Then the considerations discussed in lecture 15 apply. Otherwise we are left with a complicated problem with the following properties:

1. It has great practical importance.
2. There is a great deal of practical knowledge specific to particular parts of the problem.
3. The practical knowledge usually applies only to a single reproduction process.
4. There is no general solution, and no obvious methodology that leads to one.

It goes well beyond the scope of these notes to try to say anything comprehensive about the colour reproduction. We will make a few comments on different methods which might be tried to attack it from a perceptual point of view and give a few practical rules of thumb taken from colour reproduction experience.

16.2.1 Physical Reproduction

One approach to colour reproduction problem is to make an exact physical reproduction of the image. This is done in printing, for example, where one copy of a book or magazine is physically identical to another. This method assures that a reproduction will be totally faithful. Its advantages are:

1. From any perspective, under any illuminant the reproduction is the same.
2. Not only colour, but surface texture, gloss, etc are also reproduced exactly.
3. All observers, regardless of variability, see the reproduction as being identical to the original.

Unfortunately there are very serious disadvantages:

1. Reproduction across media is impossible: different media produce the same colour using different distributions of photons, or using different relative reflectances; different media have different surface textures; and so on.
2. Some media, such as printing, are well suited to physically identical duplication. Others, such as CRT imagery or photography, are less well suited.
3. For media which are not well suited costs can be incredible.
4. Exact physical reproduction is not really the issue. Most images start off as representations of something else. To be exact physical reproductions they would have to be indistinguishable from the original. But no one wants the picture of a mountain to be indistinguishable from a mountain. That misses the point of an having an image rather than the real thing.

The disadvantages mean either that you begin with a reproduction medium which is well suited to physical reproduction in which case you produce physically identical copies without much effort at all, or that you begin with a medium which is unsuited to it in which case it isn't even worth trying. If the reproduction, as is usually the case in computer graphics, is to go from one medium, the CRT, to another, slides, printing, etc, then physical reproduction is impossible.

16.2.2 Perceptual Reproduction

Somewhat less demanding than exact physical reproduction is exact perceptual reproduction. To do exact perceptual reproduction we think of colour, and demand that each point in the reproduction emit, not the same spectral power distribution, but any spectral power distribution that has the same tristimulus values as the corresponding point in the original. The advantages are clear:

1. The reproduction stimulates the viewer's visual system in exactly the same way that the original does.
2. All normal observers see the reproduction as identical to the original.

Unfortunately, there are significant disadvantages:

1. The first pertains to surface texture and like properties. Should they be produced exactly, at least to perceptual standards? If they should many of the problems of exact physical reproduction also exist for exact perceptual reproduction. In particular, reproduction from a medium to a dissimilar one is virtually impossible. If they should not then it will probably be necessary to adjust tristimulus values to compensate for differences in non-colour properties between the media. Thus, exact copies of tristimulus values might well produce a worse reproduction than some other criterion would.
2. Often different media have different colour gamuts. Then there may be tristimulus values in an original that are impossible to obtain in the reproduction.
3. In many media it is quite impossible to control the tristimulus values of each point independently of the other points.
4. The cost of such reproduction can be almost as high as for exact physical reproduction.
5. For reflective media it is impossible to control the illumination in which the reproduction will be viewed. Thus, the tristimulus values cannot be controlled. For CRT's it is impossible to control the veiling light present when they are viewed, so that it is equally impossible to control the tristimulus values of the reproduction.
6. Colours change their appearance as the size of the area they cover changes. For exact perceptual reproduction there can be no change in size, unless tristimulus values are allowed to change in compensation.

We usually begin by thinking that exact perceptual reproduction is the best reproduction method. Some experience trying to achieve it usually makes us realize that a less precise approach may be not only less expensive feasible but also perceptually more faithful. A little reflection on the nature of reproduction is likely to provide the insight that we want to reproduce the important relationships in the original, not to make something which is indistinguishable from the original. Thus we will look elsewhere for an acceptable solution to reproduction problems.

16.2.3 Acceptable perceptual reproduction

To see how to get an acceptable perceptual reproduction you need to think of reproduction across two quite different media, for example, from a CRT to a silk-screened T-shirt. The media differ in texture, colour gamut, and size; one is reflective; the other emits light; and so on. Yet it is quite possible to say: 'This reproduction is good; the other one is bad.' On what is this judgment based? Unfortunately, only the vaguest of answers can be given, along the lines that there are important relationships in the original image which are captured in the reproduction, and that the obvious distortions occur only to unimportant aspects of the image. But it is beyond our present understanding of how images are perceived to be able to specify what they are. We can only mention that this is the outstanding unsolved problem of colour reproduction, and of colour difference in general. It would be worth a lot to have an algorithm which, when applied to two images, could tell us how different an ordinary observer would judge them to be. That algorithm could provide the basis for a completely automated colour reproduction industry.

Despite having little concrete to say about the form of such an algorithm, there are a few comments which are worth making:

1. There are people who know how to do acceptable colour reproduction. They are the workers who run the successful reproduction industry that exists in North America. But this knowledge is unconscious. Can it be made explicit and systematic?

2. Colour relationships within an image are almost certainly more important than the absolute colours of any part of the image.
3. The overall luminance of an image is one of the least important aspects to get right in colour reproduction.
4. A successful algorithm for colour reproduction will likely have to know more about the objects represented in the image than it does about the pixels which make it up.

Finally here are a few rules of thumb for colour reproduction taken from the recent work by Stone et al. (1986).

Rules of thumb

Printers and photographers have been in the business of making acceptable colour reproductions for many years. They succeed in doing so by means of considerable unconscious knowledge, some of which is encoded in several rules of thumb. Here is a set of those rules. The set was successfully used by Stone et al. (1986) in a successful project to do digital colour reproduction.

1. Make the darkest achromatic colour in the original the darkest achromatic colour in the reproduction medium. In this rule, and in the following ones, 'achromatic' means perceptually achromatic under the viewing conditions (adaptation, illumination, surround, etc) in which the original (reproduction) is normally viewed.
2. Make the lightest achromatic colour in the original the lightest achromatic colour in the reproduction medium.
3. Ensure that all achromatic colours in the original remain achromatic in the reproduction.
4. Ensure that all coloured regions remain in the same colour category in the reproduction that they occupied in the original.
5. Make colours in the reproduction as saturated as the reproduction medium allows.
6. The brightest highlights, if they are achromatic, should be clipped to the edge of the colour gamut of the reproduction medium.
7. Slight rotations of the original gamut can be performed if they allow a better fit between irregularly shaped original and reproduction gamuts. Such rotations allow better saturation and better contrast in the reproduction.

16.2.4 Information Reproduction

At the lowest level of reproduction all we are trying to reproduce is the information in the image. For example, printed graphics might use black on white for text, with a single colour for highlights. Here the differentiation between black and the colour provides information to the reader and the object of reproduction is only to preserve that information. It is not important that the exact colour be reproduced, but only that the highlight function of the colour with respect to black be maintained. We could, for example, even change the highlight colour from green to blue without loss of information in many instances. What is distinctive about these situations is:

1. There is usually a limited number of colours in the image.
2. The information conveyed by the colours is known, at least in the sense of 'all these areas are the same colour; those areas are in a contrasting colour; and so on'.
3. The absolute colour values do not convey precise colour information.

This type of reproduction is quite easy to handle. Non-aesthetic results can be achieved by simple table correspondences, though the reproduction is usually at least as good as the original in this respect. The important fact is that the tabular transposition of one colour to another which is produced is quite enough to retain all the information which was encoded in the original. Thus, the reproduction is non-degrading, and adequate.

It might be said that all reproduction, particularly acceptable reproduction, is informational in nature. Thus, all we are trying to do when we make a copy of 'The Last Supper' is to get all the information which is in the original into the reproduction. Such an argument breaks down the distinction between acceptable perceptual reproduction and information reproduction. But this does not solve any reproduction problems, since the solutions which are so easy to find for the problems of information reproduction depend explicitly on knowing exactly how the colour carries information in the original. This is not known for an image like

'The Last Supper', and is not likely to be known for very many of the images which would come under the heading of perceptual reproduction. Those for which it is true can indeed be handled by informational methods.

16.3 References

Stone, M., Cowan, W. B., and Beatty, J. C., 1986, 'Color Gamut Mapping and the Printing of Digital Color Images,' in preparation for *ACM Transactions on Graphics*.

Notes

Lecture 17: A.R. Robertson

APPLIED COLORIMETRY

17.1 Introduction

There are a number of colour phenomena which, although not fully understood by colour scientists, are dealt with daily by industrial colorimetrists. To help these colorimetrists the CIE has defined a number of indices to provide a standard way of quantifying the phenomena. Undoubtedly, as knowledge of colour science increases, better indices will be developed but in the meantime the recommendations of the CIE provide a common language for communication.

17.2 Index of Metamerism

Metameric colour stimuli (or metamers) are colour stimuli of different spectral power distributions that produce the same colour under the same viewing conditions. When referring to object colours, the term 'metamers' is often used to describe the objects themselves (or rather their spectral reflectance factors) instead of the stimuli. In this case metamers are objects that, when illuminated by a given reference illuminant and viewed by a given reference observer, reflect stimuli of different spectral power distributions that produce the same colour under the same viewing conditions.

An example of the spectral reflectance factors of a pair of samples that are metameric under CIE standard illuminant D_{65} is given in figure 17.1.

If two objects are metameric under a given illuminant they will usually not be metameric (i.e. will not match) under another illuminant. The degree of mismatch is a useful measure of the degree of the original metamerism. The CIE has recognized this in its Special Metamerism Index: Change in Illuminant (CIE 1986a). It recommends that for two specimens whose corresponding tristimulus values are identical with respect to a reference illuminant and observer, the metamerism index M_t be set equal to the index of colour difference, ΔE, between the two specimens, computed for test illuminant t. The preferred reference illuminant is CIE standard illuminant D_{65}. The preferred test illuminant is CIE standard illuminant A or one of a series of F-illuminants that represent typical fluorescent lamps.

The concepts of metamerism and metamerism indices are discussed by Billmeyer and Saltzman (1981), Judd and Wyszecki (1975) and Wyszecki and Stiles (1982). Robertson (1983) has reviewed various concepts and definitions, correct and incorrect, that are prevalent in the field.

17.3 Colour Constancy

Colour constancy is the independence of a perceived object colour to changes in the spectral power distribution of the illuminant. It is a property of a single object and should never be confused with metamerism which is the property of a pair of objects. The only link is that at least one member of a metameric pair must exhibit a lack of colour constancy.

To quantify colour constancy, the CIE (1986b) has introduced the concept of a corresponding colour. This is a specimen with a set of tristimulus values which, under a specified condition of viewing, will elicit the same colour appearance as some other specimen with a different set of tristimulus values does when specimens of the same size and shape are viewed under different illuminants. A series of equations for predicting corresponding colours is proposed for testing.

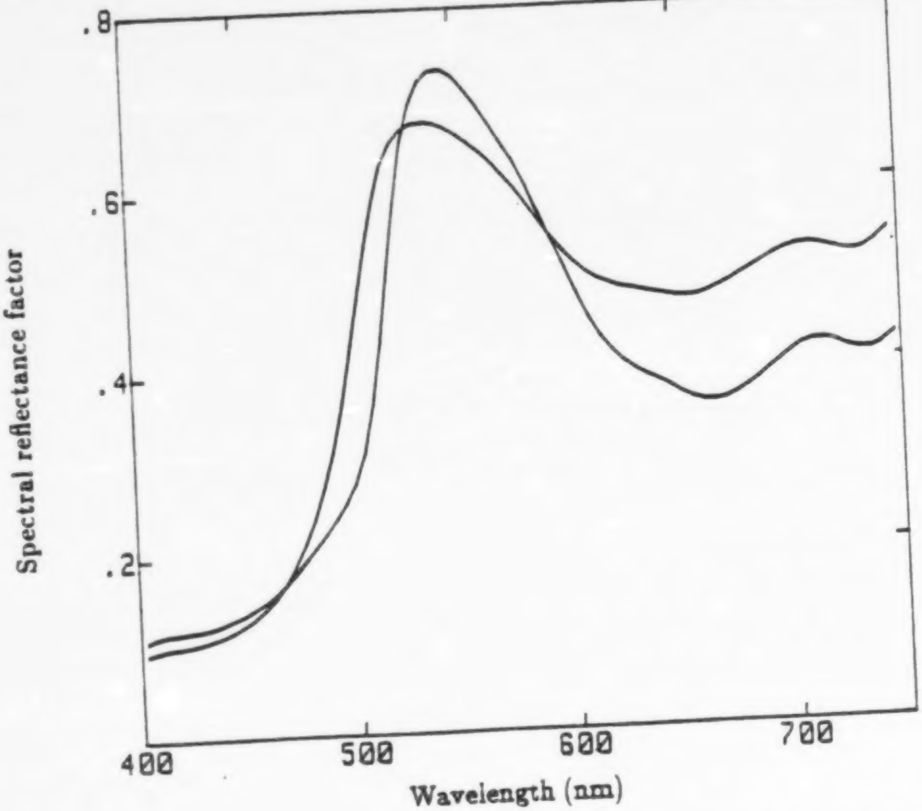

Figure 17.1. Spectral reflectance factors of a pair of yellow-green samples that are metameric under D_{65}.

17.4 Colour Rendering

Colour rendering is a general expression for the effect of a light source on the colour appearance of objects in conscious or subconscious comparison with their colour appearance under a reference light source.

The CIE (1974) has recommended a test colour method for specifying colour rendering. It rates lamps in terms of a colour rendering index, R, that is based on the colour difference ΔE of a test object under the test lamp in comparison with its colour under a reference lamp. A correction is made for chromatic adaptation to the colour of the source if the test and reference source colours are different. The index is given by $100 - 4.6\Delta E$. The most commonly quoted index is the CIE General Colour Rendering Index, R_a, which is the mean of the indices for eight defined test-colour samples.

A good discussion of colour rendering and the CIE index is in Kaufman (1984). Figure 17.2 is taken from this reference and illustrates the basis of the CIE general index. Note that the chromaticity diagram used is the (u, v) diagram which is a forerunner of the (u', v') diagram on which the CIELUV colour space is based.

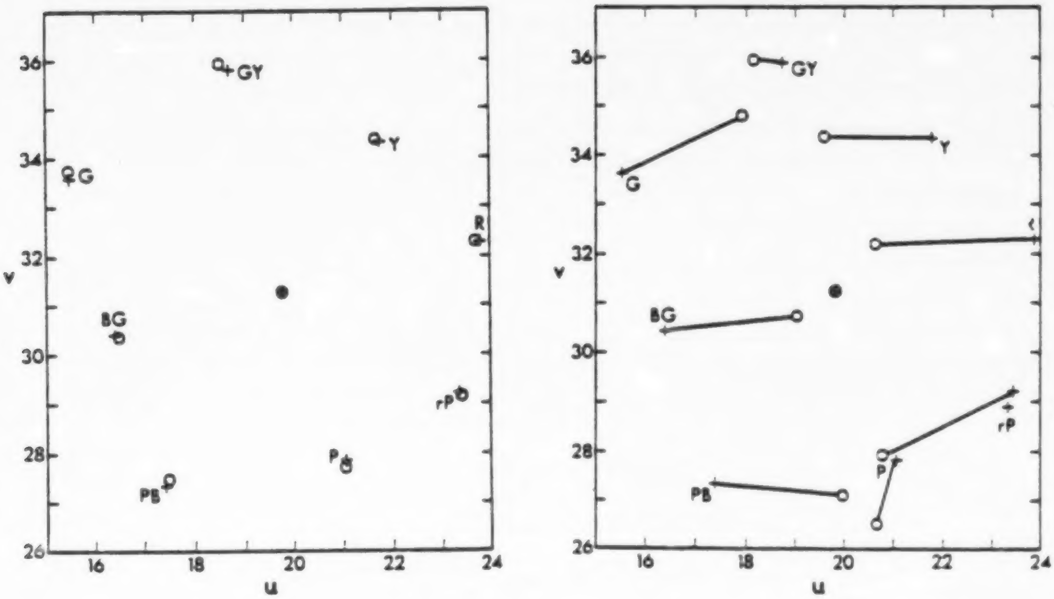

Figure 17.2. Basis for the CIE general colour rendering index in the CIE 1960 uniform chromaticity diagram. For test lamp A (left) the index is 95 with each of the test colours (0) almost the same as the standard (+). For test lamp B (right) the index is 18 with each of the test colours (0) a long way from its colour under the standard (+).

17.5 Whiteness

The evaluation of whiteness is important in several industries (notably the textile, paper and detergent industries). To promote uniformity of practice the CIE (1986a) recommends the following formula for whiteness.

$$W = Y + 800(x_n - x) + 1700(y_n - y).$$

The chromaticity coordinates x_n, y_n are those of the illuminant.

In conjunction with this recommendation, the CIE recommends the following formula for 'tint' to measure the degree of reddishness or greenishness in whites.

$$T = 1000(x_n - x) - 650(y_n - y)$$

When the formulae are used with the 1964 10° standard observer, the whiteness formula is unchanged but the tint formula becomes

$$T_w = 900(x_n - x) - 650(y_n - y)$$

17.6 References

Billmeyer, F.W., Jr. and Saltzman, M., 1981, *Principles of Color Technology*, New York: Wiley.

CIE, 1974, *Publication No. 13.2. Method of Measuring and Specifying Colour Rendering Properties of Light Sources*, Paris: Commission Internationale de l'Eclairage.

CIE, 1986a, *Publication No. 15.2, Colorimetry*, Paris: Commission Internationale de l'Eclairage.

CIE, 1986b, 'Method for predicting corresponding colors with a change in chromatic adaptation to illumination proposed for testing', *CIE Journal*, to be published.

Judd, D.B. and Wyszecki, G., 1975, *Color in Business, Science and Industry*, New York: Wiley.

Kaufman, J.E., 1984, *IES Lighting Handbook - Reference Volume*, New York: Illuminating Engineering Society of North America.

Robertson, A.R., 1983, 'Critical Review of Definitions of Metamerism', *Color Res. Appl.*, **8**: 189-191.

Wyszecki, G. and Stiles, W.S., 1982, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, New York: Wiley.